BIOMETRIKA

BIOMETRIKA

CONTENTS OF VOLUME 42

Memoirs and Miscellanea

Contents

Book Reviews

Contents

BIOMETRIKA PUBLICATIONS: BOOKS OF TABLES

Issued by the Cambridge University Press, Bentley House, London, N.W.1
and obtainable from any bookseller

Tables of the Incomplete B-Function
EDITED BY KARL PEARSON
59 pages of Introduction and 494 pages of Tables

Price: 55s. net

Tables of the Incomplete Γ-Function
EDITED BY KARL PEARSON
31 pages of Introduction and 164 pages of Tables

Price: 42s. net

Tables of the Complete and Incomplete Elliptic Integrals
(from LEGENDRE'S *Traité des Fonctions Elliptiques*. With autographed portrait of LEGENDRE)
39 pages of Introduction by KARL PEARSON and 94 pages of Tables

Price: 12s. 6d. net

Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples
BY F. N. DAVID
38 pages of Introduction, 55 pages of Tables, 10 Diagrams and 4 Charts

Price: 17s. 6d. net

NEW PUBLICATION NOW AVAILABLE
Biometrika Tables for Statisticians, Vol. I

The two volumes of *Tables for Statisticians and Biometricians* are now out of print and will not be re-issued.

At the request of the Biometrika Trustees a complete recasting of these Tables has been undertaken by Professor E. S. PEARSON and Dr H. O. HARTLEY. Many of the old tables have been set aside or modified, tables which have been published during the last fifteen years in *Biometrika* are reproduced and some new tables added.

Volume I of the new series, which includes the statistical and auxiliary mathematical tables in more common use is now available. It contains an Introduction and 54 tables covering in all 238 pp.

Price: 25s. net

NEW STATISTICAL TABLES: SEPARATES RE-ISSUED FROM BIOMETRIKA

NEW STATISTICAL TABLES: continued

ELEMENTARY STATISTICAL EXERCISES, PART 1

These exercises are based on the numerical class work which accompanies the first year lectures to undergraduate students taking their B.Sc. degree in Statistics of the University of London. They cover the more elementary univariate theory, regression and correlation, χ^2 and significance tests, sampling distributions and quality control.

The book contains 91 pp. and is reproduced from typescript by a photographic process.

Price: 6s. 6d.

CONTENTS

STUDIES IN THE HISTORY OF PROBABILITY AND STATISTICS*

I. DICING AND GAMING (A NOTE ON THE HISTORY OF PROBABILITY)

By F. N. DAVID

University College, London

'See, this is new? It hath been already of old time.' (Ecclesiastes i. 10.)

1. A cynical archaeologist remarked recently that a symptom of decadence in a civilization is when men become interested in their own history, and he added that in the unlikely eventuality of any proof being required of the decadence of this phase of *Homo sapiens* it could be found in the present-day interest in archaeology. Most generalizations of a sweeping character such as this are unacceptable, chiefly because there is no way of putting them to proof; but the present interest of scientists in general, and of statisticians in particular, in the origins of scientific thought, is far from implying the decadence of science, whatever may be implied by an interest in the arts.

It is inviting, and at the same time profitless, to speculate why modern scientists have such an interest. The possibility of deciding priority of discovery which concerned the Victorian scientist so closely does not cause much controversy to-day, for the modern scientist would hold that to ascribe any discovery in the field of science to any single person is unrealistic. Thus, while we are taught at school, for example, that Newton and Leibnitz separately and independently 'discovered' the differential calculus, it would perhaps be more appropriate to say that Newton and Leibnitz each supplied the last link in the chain of reasoning which gave us the differential calculus—a chain which can be traced back through Pierre Fermat, Barrow, Torricelli and Galileo, and that it is surprising that there were only two mathematicians who did this.

Mathematics is essentially an expression of thought in which we build on the mental effort of our forerunners, and probability is no exception to this general rule. The real difficulty we meet with in trying to trace probability back to its origins is that it started essentially as an empirical science and developed only lately on the mathematical side. It is hard to say where in time the change came from empiricism to mathematical formalism as it appears to have taken place over hundreds of years; and the claims put forward for Pascal and Fermat as the creators of probability theory cannot entirely be substantiated.

2. When man first started to play games of chance is a time problem we shall never clearly resolve. We may place on record that it is a commonplace thing for archaeologists to find a preponderance of astragali† among the bones of animals dug up on prehistoric sites. One archaeologist stated that he had found up to seven times as many as any other bone, another put the figure at 500 (sic!), while yet a third, refusing to be drawn to a figure, stated that they were many. This fact has probably little significance. The astragalus has little marrow in it and was possibly not worth cracking for the sake of its contents as were the long bones; it is knobbly and presents no flat curves for drawing as does the shoulder

---

* [*Editorial note.* It is hoped to publish articles by a number of different authors under this general heading.]

† The *astragalus* is a small bone in the ankle, immediately under the *talus* or heel-bone. See Pl. 2*a*.

x

blade for example. All we may do is to place on record that round about 40,000 years ago there were large numbers of the astragali of sheep, goat and deer lying about.

The astragali of animals with hooves are different from those with feet such as man, dog and cat. From the comparison in Text-fig. 1 we note how in the case of the dog the astragalus is developed on one side to allow for the support of the bones of the feet. The astragalus of the hooved animal is almost symmetrical about a longitudinal axis and it is a pleasant toy to play with. In France and Greece children still play games with them in the streets, and it is possible to buy pieces of metal fashioned into an idealized shape but still recognizable as astragali.

Sheep   **Dog**
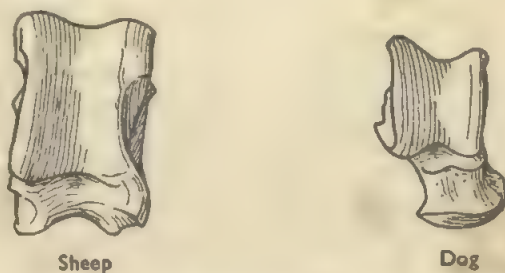
Text-fig. 1. Drawings of the astragalus in sheep and dog, natural size.

3. Some time between prehistoric man of four hundred centuries ago and the beginning of the third millennium before Christ *Homo sapiens* invented games and among these games, games of chance. We know from paintings, terra-cotta groups, etc., that the astragalus was used in Greece like the ancient quoit,* but there is no doubt from paintings on tombs in Egypt and excavated material that the use of the astragalus in games where it is desired to move counters was well established by the time of the First Dynasty. In one painting a nobleman, shown playing a game in his after-life, delicately poises an astragalus on his finger tip, a board with 'men' in front of him. A typical game of *c.* 1800 B.C. is that of 'Hounds and Jackals' illustrated in Pl. 1. The game seems similar to our present-day 'Snakes and Ladders'; the hounds and jackals were moved according to some rule by throwing the astragali found with the game and shown in the figure. Variants of this game were undoubtedly played from the time of the First Dynasty (*c.* 3500 B.C.).

It is possible but not altogether likely that these games originated in Egypt. They certainly did not originate in Greece, as has been claimed for reasons which we shall give later. However, Herodotus, the first Greek historian, like his present-day counterparts, was willing to believe that the Greeks (or allied peoples) had invented nearly everything. His claim that the Lydians introduced coinage has about as much foundation as his claim regarding games of chance. He writes (*c.* 500 B.C.) about the famine in Lydia (which was *c.* 1500 B.C.) as follows:

The Lydians have very nearly the same customs as the Greeks. They were the first nation to introduce the use of gold and silver coins and the first to sell goods by retail. They claim also the invention of all games which are common to them with the Greeks. These they declare that they invented about the time that they colonized Tyrrhenia, an event of which they give the following account. In the days of Atys, the son of Manes, there was great scarcity through the whole land of Lydia. For some time the Lydians bore the affliction patiently, but finding that it did not pass away,

* From the name 'knucklebone' we might infer that among the early games were those in which the astragali were balanced on the bones of the knuckles and then tossed and caught again.

Plate 1

The game of Hounds and Jackals.

David: *Studies in the History of Probability and Statistics*

(a)

(b)

(c)
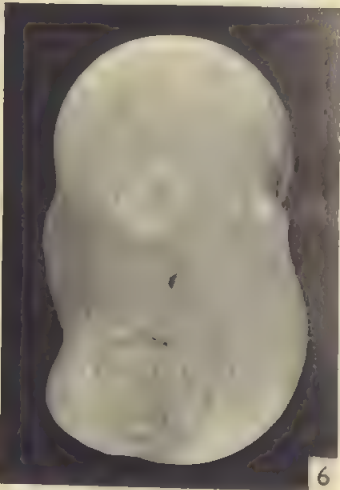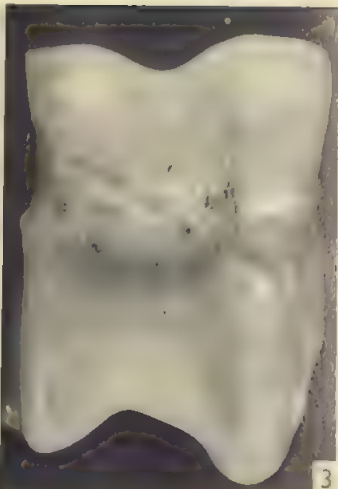
they set to work to devise remedies for the evil. Various expedients were discovered by various persons; dice and knuckle-bones (i.e. astragali) and ball and all such games were invented, except tables,* the invention of which they do not claim as theirs. The plan adopted against famine was to engage in games on one day so entirely as not to feel any craving for food, and the next day to eat and abstain from games. In this way they passed eighteen years.

In yet another commentary we are told that games of chance were invented during the Trojan war by Palamedes. During the 10 year investment of the city of Troy various games were invented to prevent the soldiers' morale suffering from boredom.

4. The game of ball is mentioned by Homer and according to Plato was evolved in Egypt. It is not, however, a game of chance. The story of dice we shall return to, but we may first carry the story of the astragalus a little further. In the early part of the first millennium it would seem that astragali were used by both adults and children for their leisure games. Homer (c. 900 B.C.) tells us that when Patroclus was a small boy he became so angry with his opponent while playing a game of knucklebones that he nearly killed him. Another writer of the same period tells us that students played knucklebones everywhere, that they were acclaimed as presents and that as a prize for handwriting one student was given eighty astragali all at once! It is not difficult to imagine the small boys of that era collecting astragali as they collected marbles, much as the boys of our own era still do.

That the astragalus was used commonly in the gaming which the Greeks and later the Romans conducted with zeal and passion, the references in the literature of that time leave no room for doubt. One of the chief games may have been the simple one of throwing four astragali together and noting which sides fell uppermost. The astragalus has only four sides on which it may rest, and it has been deduced, among others by Nicolas Leonicus Thomeus (1456–1531), that a common method of enumeration was that the upper side, broad and slightly convex counted 4, the lower side broad and hollowed 3, the lateral side narrow and flat 1 and the other lateral which is slightly hollow 6. These aspects of a sheep's astragalus are shown in Pl. 2a. (With present-day astragali the probabilities of scoring 1 and 6 are each approximately equal to 1/10 and those of 3 and 4 approximately 4/10.) The worst throw for the Greeks with one bone was unity which they called the dog, and sometimes the vulture. The best of all throws with four knucklebones was the throw of Venus when all four sides were different which has an actual probability of about 1/26. But at different times and in different games the numbers must have been varied, for the throw of Euripides with four astragali, discussed by several fifteenth-century writers, was worth 40. How the bones fell to achieve this result is not stated, although Cardano writing in the sixteenth-century states that it was four fours. (Probability c. 1/39.)

5. In classical Rome the astragalus was imitated in carved stone with figures and scenes incised on the sides. A typical example is illustrated in Pl. 2b. Stone astragali have also been found in Egypt. At this time too we have the production of lewd figures in metal or bone varying in size from about 1 cm. to over 1 in. in height. That these figures were used for gaming may be deduced from the fact that the six possible positions in which the figure may fall are each marked with a number of dots.†

Besides the astragali it appears possible that throwing sticks were also used for games of chance, although it may be that they had a greater religious significance; we shall return

* This may have been an early form of backgammon or may have been shuffle-board.
† I have not tested these figures for bias. They are a development, I think, of dice rather than astragali.

to this point later. The throwing stick was made of wood or ivory and was often approximately 3 in. in length with cross-section when square of about 1 cm. each way. Such throwing sticks were known to the ancient Britons, to the Greeks, the Romans, the Egyptians and to the Maya Indians of the American continent. Sometimes the sticks are elliptical in cross-section with major axis of approximately 1 cm., but they are all alike in having only four numbers on them, one at each end of the upper face and one at each end of the lower parallel face. In the European throwing sticks the majority of numbers are marked by small engraved circles, but they are sometimes indicated by cuts in the wood or ivory. The Maya throwing sticks are marked by coloured scratches in ivory. The actual numbers marked vary. They are mostly 1, 2, 5, 6, but 3 and 4 have also been noticed. These throwing sticks are of little importance in gaming. They are mentioned because it is interesting to note the likelihood that gaming originated at many points, and, although this is a remark one could not defend, that it possibly was originally a debasement of a religious ceremony.

6. The six-sided die may have been obtained from the astragalus by grinding it down until it formed a rough cube. The Musée de Louvre has several astragali which have been treated in this way, but one cannot imagine they formed very satisfactory dice. The honeycombed (or cancellous) bone tissue has been exposed in several places and the crude die would clearly not have a long life. Whether the die was evolved in this way or not the evolution must have taken place some considerable time before Christ. The earliest die known was excavated in northern Iraq and is dated at the beginning of the third millennium. It is described as being of well-fired buff pottery. The dots are arranged as shown in Text-fig. 2(i), the edges at A and B being imagined folded away from the reader. It will be noted that the opposite points are in consecutive order, 2 opposite 3, 4 opposite 5 and 6 opposite 1.

A die excavated in Mohenjo-Daro (Ancient India) is also dated as third millennium, and it is also made of hard buff pottery. The order of the points is again consecutive, but this time we have 1 opposite 2, 3 opposite 4, 5 opposite 6. Few other dice have been recorded in this millennium. At the time of the XVIIIth dynasty in Egypt (c. 1400 B.C.) a die with the markings shown in Text-fig. 2(ii) must have been in play. The arrangement of the five dots is unusual. Somewhere about this time, however, the arrangement of the numbers settled down to the familiar two-partitions of 7 opposite one another as shown in Text-fig. 2(vi), which arrangement has persisted until to-day. Out of records, collected by the present writer, of some fifty dice of classical times made of crystal, ivory, sandstone, ironstone, wood and other materials, forty had the two-partition of 7 arrangement. A twelfth-century (A.D.) Greek bishop wrote that this was the way in which a die should be marked, and a sixteenth-century gambler theorizes that this arrangement was chosen to make it easy to check whether all the numbers had been marked on the die and no figure duplicated at the expense of leaving out another. One die of the first millennium is said to have 9 opposite 6, 5 opposite 3, and 4 opposite 2. It may have been especially made for a particular game; alternatively, it is possible that it may have some ceremonial significance. This is possibly also true of a die marked as in Text-fig. 2(iii), although it might have been a die used for cheating.

7. That dice were used in Egypt is clear from the XVIIIth-dynasty specimen. It is thought, however, that dicing did not become common until the advent of the Ptolemaic dynasty (300 to 30 B.C.) which originated from Greece. Several dice are known of this

period including a beautiful specimen in hard brown limestone of side c. 1 in., which has the sacred symbols of Osiris, Horus, Isis, Nebhat, Hathor and Hesterhet engraved on the six sides. This would almost undoubtedly have been used for some form of divination rites.

(i)

(ii)

(iii)

(iv)

(v)

(vi) Rock Crystal

(vii) Iron

(viii) Marble
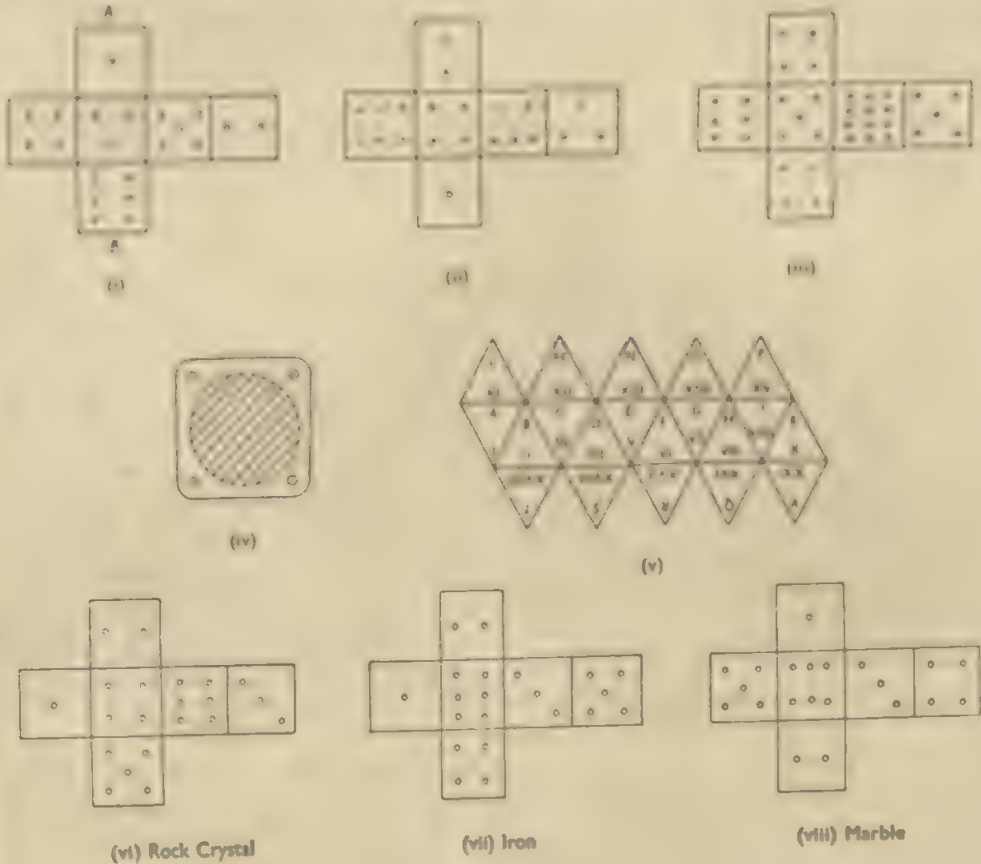
Text-fig. 2

Dice in Britain were in a very primitive state at the time of Christ. The pieces used then were formed by roughly squaring the long bone of an animal and cutting it into sections to form objects approximately cubical in shape. The marrow was taken out leaving a hollow square cylinder of which the cross-section in diagrammatic form is sketched in Text-fig. 2 (iv). (The British Museum has two of these.) These primitive dice had the two partition of 7 arrangement, 3 being opposite 4 on the hollow ends. Some dice had a 3 on each end and no 4. Several dice of this kind have been excavated in the chalk and flint country and dated late in the first millennium.

The working out of the geometry of solid figures by Greek mathematicians appears to have been followed almost immediately by the construction of polyhedral dice. A beautiful icosahedron in rock crystal now at the Musée de Louvre is the most famous of these. (In the diagram of Text-fig. 2 (v) it may be imagined the outline is folded away from the reader.) A figure with 19 faces badly cut but apparently imagined to be rectangular has a roman digit on each face from I to X, and above that the numbers rise by tens to C. The

number LXXX is missing, but the number XX appears twice on one face. There was also a die of 18 faces probably formed by beating out a cubical die, a die with 14 faces and so on.

Faked dice were also not unknown. Apart from the device of leaving out one number and duplicating another it is stated that hollow dice have been found dating from Roman times. The unity on the face of the die forms a small round plate which can be lifted. It is suggested that a small ball of leather could be crammed into the hollow of the die through this hole in such a way as to cause the die to tend to fall in a predetermined manner.

8. Gaming reached such popularity with the Romans that it was found necessary to promulgate laws forbidding it except at certain seasons. What game was played by the common people we do not know, but there are many references to those played by the emperors. In Suetonius's *Life of Augustus* (Loeb's translation) we find:

> He (Augustus) did not in the least shrink from a reputation for gaming and played frankly and openly for recreation, even when he was well on in years, not only in the month of December, but on other holidays as well and on working days too. There is no question about this, for in a letter in his own handwriting he says, 'I dined, dear Tiberius, with the same company; ...We gambled like old men during the meal both yesterday and to-day, for when the dice were thrown whoever turned up the "dog" or the 6, put a denarius in the pool for each one of the dice, and the whole was taken by anyone who threw the Venus'.

There are several other references to gaming in this *Life*. Whether the word talis should be translated as dice or astragali (knucklebones) is a moot point. The die as we know it is usually referred to as 'tessera'. The astragalus is often called the talus (or heel-bone), and this is the word Suetonius actually used. From the description of the play it would seem appropriate to read knucklebones for dice.

In Suetonius's *Life of Claudius* we are told that Claudius was so devoted to dicing that he wrote a book about it, and that he used to play while driving, throwing on to a board fitted especially in his carriage. From another source we learn that he played right hand against left hand.

9. These two instances are chosen to illustrate the passion for gaming which apparently possessed the Romans, and it is possible to cite many others. The question which constantly recurs to one while studying these games of the past is 'Why did not someone notice the equi-proportionality property of the fall of the die?' It is understandable that no theory was made to describe the fall of the astragalus. But the Greeks had performed the necessary abstraction of thought to make the mathematical idealization of the cube (and other solid figures); at first sight it seems curious that mathematicians did not then go on a little further and give equal weight to each side of the cube and so on. For if dicing and gaming generally were carried on by so many persons for so long that it was thought necessary to prohibit them, surely someone must have noticed that with a cube on the average any one side turned up as frequently as any other? We can only make guesses on this point, but it would seem to the writer that there are two possible explanations, the imperfections of the dice and their use in religious ceremonies.

10. *Imperfect dice.* We speak of a true or a fair die nowadays when we mean that there is no bias apparent when the die is thrown. In Roman times, and presumably earlier, it seems to have been the exception rather than the rule for the die to be true. Many dice of the classical period have been thrown by the writer and they were nearly all biased but not all in the same way. For example, three classical dice from the British Museum gave the results shown in the table from 204 throws each. The arrangement of the pips on the dice

were as in Text-fig. 2 (vi), (vii) and (viii). The rock crystal is a beautifully made die; the others are a little primitive, and the sides of the iron die are only approximately parallel. The marble and the iron dice are obviously biased, and this was true of many of the other dice examined. A photograph of a wooden die of the classical period is given in Pl. 2(c). It will be noted that one of the faces shown is not square, and the impression one has is that the owner picked up a piece of wood of convenient shape, smoothed it a little and engraved the pips. It would therefore have been difficult, except over a long period, to notice any regularity.

Number of pips ...	1	2	3	4	5	6
Rock crystal	30	38	31	34	34	37
Iron	35	39	30	21	37	42
Marble	27	28	23	47	25	54

11. *Divination.* In spite of the imperfections of the dice it is probable that some theory might have been made if magic or religion or both had not been involved also. A scheme whereby the deity consulted is given an opportunity of expressing his wishes appears to be a fundamental in the development of all religions. As late as 1737 we have John Wesley deciding by the drawing of lots whether to marry or not (John Wesley's *Journal*, vol. 1, 1737, Friday, 4 March), and in the practices of present-day primitive tribes we get an echo from the classical era. At that time pebbles of diverse shapes and colours, arrows, astragali and dice were all used to probe the divine intention. In the temples there were various and varied rites attached to the process of divination by lot, but the main principle was the same. The question was posed, the lot was cast, the answer of the god was deduced. The dice (astragali, etc.) were thrown sometimes on the ground, sometimes on a consecrated table.

It was customary in classical Greece and Rome for the four astragali of the gamblers to be used in the temples. The prediction was that the throw of Venus (1, 3, 4, 6 uppermost) was favourable and the dogs unfavourable. In the temple of the oracles tablets were hung up and the priest, or possibly the suppliant, interpreted the throw of the four bones by reference to the tablets. Cases have been recorded, however, where five astragali were used. Greek inscriptions found in Asia Minor give a fairly complete record of how the throws of five were interpreted. Each throw was given the name of a god. Thus Sir James Frazer translates (commentary on Pausanias):

1. 3. 3. 4. 4 = 15   *The throw of Saviour Zeus*
One one, two threes, two fours,
The deed which thou meditatest, go, do it boldly.
Put thy hand to it. The gods have given these favourable omens.
Shrink not from it in thy mind. For no evil shall befall thee.

It is not clear whether the order of the numbers is important or not. If order does not matter then the probability of this throw is about 0·08.* The tesserae of the gambler were also used in divination ceremonies as well as the astragali, and it is possible that the same interpretation was given to the numbers falling uppermost, although the presence of the 2 and the 5 would make this a little awkward.

* I propose to write at greater length on 'divination probabilities' on a further occasion.

In addition to the divination carried out by the priests it was apparently a commonplace for individuals to perform acts of divination with regard to events in their daily lives. Thus Lucian, telling the story of the young man who fell madly in love with Praxiteles's Venus of Cnidos, writes:

> He threw four knucklebones on to the table and committed his hopes to the throw. If he threw well, particularly if he obtained the image of the goddess herself, no two showing the same number, he adored the goddess, and was in high hopes of gratifying his passion: if he threw badly, as usually happens, and got an unlucky combination, he called down imprecations on all Cnidos, and was as much overcome by grief as if he had suffered some personal loss.

Again we have from Propertius:

> When I was seeking Venus (i.e. good fortune) with favourable tali, the damned dogs always leaped out.

12. It is perhaps of interest here to interpolate a note on divination as reported practised by the Buddhists of present-day Tibet. According to Hastings (*Dictionary of Comparative Religions*) the simplest method is carried out by the people themselves. Many laymen are equipped with a pocket divination manual (*mô-pe*) and the augury found by casting lots. This lot-casting can either be odds and evens (the random pouring of grain, pebbles or coins from a horn, cup, etc.) or dice on a sacred board or cards on which there are magic signs, or sheets or passages of scriptures drawn from a bowl. The reincarnation prediction is, it is said by Waddell,* usually carried out by a priest. The rebirth chart seen by the writer consists of 56 2 in. squares (8×7). Each square corresponds to a future state. A six-sided die with letters on it is thrown down on to the rebirth chart, and according to the square on which it lands and the letter which falls uppermost so the priest predicts. Waddell, who visited Tibet as a member of a British Mission, obtained one of these charts and a die (*c*. 1893). He remarks: 'The dice (sic!) accompanying my board seems to have been loaded so as to show up the letter Y, which gives a ghostly existence, and thus necessitates the performance of many expensive rites to counteract so undesirable a fate.' Possibly a similar chicanery was practised in Roman times! It would seem a reasonable inference anyway that the mystery and awe which the religious ceremony would lend to the casting of lots for purposes of divination would prevent the thinking person from speculating too deeply about it. Any attempt to try to forecast the result of a throw could undoubtedly be interpreted as an attempt to forecast the action of the deity concerned, and such an act of impiety might be expected to bring ill luck in its train. In addition, as we have noted, a method for such forecasting could not easily be made owing to the imperfections of most of the dice. On the other hand, it is possible that probabilities were known to the priests since the ceremonial dice are well made.

13. Through the Dark Ages the Christian church appears to have carried on guerilla warfare against gaming with knucklebones and dice. The writers of the Renaissance make many references to bishops who write *de aleatoribus* or *contra aleae ludum* during the first fourteen hundred years of the Christian era. It is likely therefore that the bishops wished to get rid of the sortilege as a religious ceremony, and they succeeded to a certain extent in doing this, although divination by lot still survives to-day in the Moravian sect. What the bishops could not do was to stop men playing games of chance. There are several references in early French literature to gaming. The play of Jean Bodel, *Le Jeu de Saint Nicolas*, written *c*. A.D. 1200, has a scene where thieves are gambling in a tavern. They are playing

---

* L. A. Waddell, *The Buddhism of Tibet*, W. Heffer and Sons Ltd. 1934 (2nd edition).

the dice game of 'Le hasard'.* the rules of which were set down clearly by Pierre-Raymond Montmort in his book some five hundred years later. (*Analys sur le jeux d'azard* 1708, p. 113). Bodel's play is interesting for the suggestion that the thieves knew how to manipulate the dice to produce a desired result.

14. With the invention of printing (*c.* 1450) and its rapid development during the latter half of the fifteenth century the references to games of chance become more numerous, but there seems to be no suggestion of the calculation of probabilities. Thus we find in the writing of François Rabelais—a man who might be expected to know the latest in games of chance as played in taverns—the following interesting passage: 'Then they studied the Art of painting or carving; or brought into use the antic play of tables, as Leonicus hath written of it or as our good friend Lascaris playeth at it'† (*Gargantua and Pantagruel*, Urquhart's translation. Book 1, Chapter xxiv). *Gargantua and Pantagruel* was issued in sections at intervals between 1532 and 1552. The date of this reference will therefore be not long after 1532.

The Leonicus of the reference is Nicolas Leonicus Tomeus, a professor of Greek and Latin at Padua who was born at Venice in 1456. He was well known for his learning and philosophical bent and acted as tutor to the English Cardinal Pole when as a young man he visited Italy. According to Erasmus he was 'a man equally respectable for the purity of his morals and the profundity of his erudition'. His letters, which have been translated by Cardinal Gasquet, give an interesting picture of the life of an intellectual of that time. He died at Padua in 1531 and his collected works were printed at Basel in 1532. Rabelais is clearly referring to Leonicus's treatise *Sannutus, sive de ludo talario,* a dialogue in the manner of Plato concerning the game of knucklebones (astragali). There is, however, little relevant to the calculus of probability in this work. The discussion turns on references to the game in Roman literature and a description and argument of the value of the various types of throw.‡

A similar type of disquisition was written by Calcagnini about this time but possibly a little later than that of Leonicus. Celio Calcagnini was born at Ferrara in 1479 and died there in 1541. He was a poet, a philosopher and astronomer of repute; his treatise *Quomodo caelum stet, terra moveatur, vel de perenni motu terrae commentatio,* in which he held that the earth moved round the sun, anticipated Galileo Galilei by some years, for Galileo was not born until 1564. The dissertation of Calcagnini entitled *De talorum, tesserarum ac calculorum ludis ex more veterum* is less philosophical in tone than that of Leonicus. It is of interest to probabilists only in that it was an influence over Cardano, who, from his several references, had clearly studied it closely.

* According to the editor, F. J. Warne, of the text of the play, *Le Jeu de Saint Nicolas,* 'hasart' meant the throw of a certain number of points at dice, varying according to the game played. In present-day probability theory the meaning is of course much wider.

† Rabelais actually wrote 'en usage l'anticque jeu des tables ainsi qu'en a escript Leonicus'. Duchat in the commentary on the 1741 edition says 'Ce n'est point *tables* qu'il faut lire ici, comme dans toutes les Editions, mais *tales*'. Presumably Duchat (followed by later commentators) makes this correction because of the work of Leonicus referred to. It is just possible that Rabelais meant what he wrote and that he was referring to the ancient board game (from which the modern game of backgammon developed) in which the 'men' may have been moved by throwing astragali, the counting of the throws being that described by Leonicus.

‡ I have not been able to trace why Lascaris is mentioned by Rabelais. Andre-Jean Lascaris surnamed Rhyndaconus (1445–1535), a Greek scholar born in Phrygia, was Librarian to François I. He rescued many Greek manuscripts from the Turks. Possibly he collected references to gaming in Greek literature much as Leonicus did for Roman?

15. We arrive at the sixteenth century, then, with a well-known humanist Leonicus, and a great astronomer Calcagnini, writing on games of chance with no attempt or reference to the calculation of a probability. (This does not mean of course that some calculations had not been made in a manuscript which we do not know about.) There were, moreover, other scholars and bishops writing on the same topic about this time, so that interest in the subject was keen. As far as we know at present it was left to Gerolamo Cardano to make the step forward. Cardano, the illegitimate son of a geometer, Fazio Cardano, was born in Pavia in 1501. His illegitimacy was a bar to his professional advancement on more than one occasion, and it is possible that the bitterness engendered by this fact was responsible for his not too scrupulous regard for the attribution of other scientists' ideas. The crime of plagiarism was a common accusation among scientific workers of the sixteenth and seventeenth centuries, but of none was it raised more loudly than of Cardano who was strongly disliked by his contemporaries and despised by his successors. Until about the middle of the nineteenth century his biographers unite in regarding him as a charlatan; possibly at the present time the pendulum has swung too far the other way, and more is read into his writings than is justified. The truth would seem to lie somewhere between the extremes of charlatan and persecuted savant.

Cardano was physician, philosopher, engineer, pure and applied mathematician, astrologer, eccentric, liar and gambler, but above all a gambler. He himself owns that on one occasion he sold his furniture and his wife's possessions in order to get money to indulge his passion for gaming, and there is no doubt that this passion was one of the things which ruled him through his whole life. His chief interest professionally was medicine, but he interested himself also in the communication of spirits and the casting of horoscopes. He does not seem to have been too successful at this last, but he was not deterred from casting that of Jesus, a performance the impiety of which probably led to his imprisonment. Even allowing for the exaggerations of his biographers there seems to be no doubt that he was eccentric to the point of madness. This did not prevent him, however, from making contributions to pure mathematics, and it is to this combination of pure mathematician and gambler that we owe the *Liber de Ludo Aleae*. This treatise was found in manuscript in Cardano's papers after his death in Rome in 1576, and was first published in his collected works in 1663 at Lyon. Cardano implies that it was written *c.* 1526; the exact date is not important since no question of priority or plagiarism is involved, but it is curious that a manuscript of this kind should have survived fifty years of his remarkably variegated career.

16. The first complete translation of *de Ludo Aleae* into English is given in *Cardano, the Gambling Scholar*, by Oystein Ore, published in 1953. Ore remarks that the book is badly composed and that understanding of Cardano's work has possibly been hindered by this. There are some, however, who will not agree with his commentary on the treatise and who may feel that as much prescience is now attributed to Cardano as there was before too little. The crux of Cardano's work is to be found in the section entitled 'On the cast of one die' in Ore's translation:

> The talus has four faces and thus also four points. But the die has six; in six casts each point should turn up once; but since some will be repeated, it follows that others will not turn up. The talus is represented as having flat surfaces, on each one of which it lies on its back;...and it does not have the form of a die. One half of the total number of faces always represents equality; thus the chances are equal that a given point will turn up in three throws, for the total circuit is completed

in six, or again that one of three given points will turn up in one throw. For example, I can as easily
throw one, three or five as two, four or six. The wagers are therefore laid in accordance with this
equality if the die is honest....

We have therefore the necessary abstraction made: if the die is honest, i.e. if we may give
equal weight to each side, then we may calculate the chances. There is no doubt, I think,
that Cardano was led to this conclusion empirically and his generalization of it is partially
wrong. For he goes on to discuss casts of two dice and three dice giving tables which are
correct if 'the dice be honest'. When we come, however, to the section 'On play with
knucklebones', it seems that he falls into error. The knucklebones (or astragali) have four
sides. The different combinations of numbers which may arise in the throwing of four
astragali are correctly enumerated, but the chances are calculated under the assumption that
all sides are equi-probable; which they are not. Possibly Cardano had never played with
astragali, for it is likely that if he had he would have noticed that to assume the sides of the
astragalus had equal weight in his enumeration of alternatives was not adequate. But this
fumbling suggests that he was not quite clear in his own mind about what he was proposing.

I do not think that the fact that Cardano did not quite see the mathematical abstraction
clearly can detract from the fact that he did, on paper at any rate, as far as we know,
calculate the first probability by theoretical argument, and in so doing he is the real begettor
of modern probability theory. The claims of his biographer that he anticipated the law of
large numbers, etc., may not be acceptable; it would appear that Cardano was judging
from his experience rather than his algebra.

17. It would be strange if Cardano, following the mode of his age, did not communicate
some of his thoughts about gaming to his pupils. Fear of being accused of plagiarism, fear
of being plagiarized, may have kept him silent, but the whole tone of his treatise is a
practical one; practical advice about playing, laying odds and so on make up a large portion
of it. He would therefore almost certainly have discussed its contents with his friends,
particularly if he thought about it over as long a period of time as he suggests. The fact that
de Ludo Aleae did not appear in print until 1663 does not therefore seem to be a reason
why Cardano's ideas should not have been common knowledge to scholars in Italy after his
death, and the way in which Galileo-Galilei plunges into *his* discussion of dice playing,
without much preamble, tends to lend colour to this.

Galileo-Galilei was born in Pisa in 1564, the son of Vincent Galilei, a musicographer well
known in his day. He died in 1642 at Arcetri after a career as full of achievement as any
that has ever been known. His contributions to science, both as astronomer and as mathe-
matician, are striking for their originality of thought and clarity of purpose. Why this
prince of scholars has never received the full recognition which is his due it is difficult to
say. It is thought by some modern writers that his sensible recantation of the earth's
movement, after physical torture at the hands of the Inquisition at the age of 70, has caused
a revulsion to him among the scientists of later years. This is probably not so; what is more
likely is that the envious fellow-scholars who delivered him to the Inquisition conspired
after his death to belittle the work which he had done. In this they were possibly helped
by Galileo's literary style which is noteworthy for clarity but not brevity, being in fact
prolix and tedious in the extreme; no *i* is left undotted, no *t* is left uncrossed.*

* E. S. Pearson suggests to me that this prolixity was one which Galileo shared with many other
Renaissance writers, and that it arose from the struggle which the early mathematicians must have
had to formulate mathematical abstractions on paper. I think that this may well be so.

This being so if there was any doubt about the general method of procedure in calculating chances with a die we should have had a long disquisition on the subject. However in *Sopra le Scoperte de i Dadi** he plunges straight away into his argument. The problem is one already touched on by Cardano. Three dice are thrown. Although there are the same number of three partitions of 9 as there are of 10, yet the probability of achieving 9 in practice is less than that of throwing 10. Why is this? I quote a little from E. H. Thorne's translation of this note. The note begins:

> The fact that in a dice game certain numbers are more advantageous than others has a very obvious reason, i.e. that some are more easily and more frequently made than others, which depends on their being able to be made up with more variety of numbers. Thus a 3 and an 18, which are throws which can only be made in one way with 3 numbers (that is, the latter with 6, 6, 6 and the former with 1, 1, 1, and in no other way) are more difficult to make than, for example, 6 or 7, which can be made up in several ways, that is a 6 with 1, 2, 3 and with 2, 2, 2 and with 1, 1, 4 and a 7 with 1, 1, 5; 1, 2, 4; 1, 3, 3; 2, 2, 3. Again, although 9 and 12 can be made up in as many ways as 10 and 11 and therefore they are usually considered as being of equal utility to these, nevertheless it is known that long observation has made dice players consider 10 and 11 to be more advantageous than 9 and 12.

This extract serves to show how he begins the topic assuming that the calculations are known; it also serves to illustrate the prolixity of Galileo's style. After some discussion of the six 3 partitions of 9 and of 10, he goes on:

> Since a die has six faces and when thrown it can equally well fall on any one of these, only six throws can be made with it, each different from all the others. But if together with the first die we threw a second, which has also six faces, we can make 36 throws each different from all the others, since each face of the first die can be combined with each of the second....

After saying that the total number of possible throws with three dice are 216, he gives a table of the number of possible throws for a total of 10, 9, 8, 7, 6, 5, 4, 3, noting that the numbers 11–18 inclusive are symmetrical with these. Thus the number of possible throws for 10 is 27 and 25 for 9. His treatment of the problem is exactly that which we should use to-day and leaves us in no doubt that the calculation of a probability from the mathematical concept of the equi-probable sides of the die was clearly known to the sixteenth-century mathematicians of Italy. We can marvel at the person asking Galileo the question: he obviously gambled sufficiently to be able to detect a difference in empirical probabilities of 1/108.‡

18. Galileo's collected works were first published in Bologna in 1656, but this fragment on gambling was not included. It does appear in the more complete collection published at Florence in 1718. Since, however, Galileo thought the problem of little interest, for he did not pursue it, there seems to be no reason why he should have made a secret of it, and following the custom of his day he probably instructed his pupils. At any rate it is evident that the mathematical probability set was no stranger to the French mathematicians of the seventeenth century, as is witnessed by the now famous correspondence between Pascal and Fermat in 1654. The first letter of the series, from Pascal to Fermat, setting out the problem of points is missing. We have, however, Fermat's reply to it, and the subsequent

---

* This is Galileo's own title. *Consideratione sopra il Giuoco dei Dadi*, a later title, appears first in the collected works of 1718.

† Like Pascal sometime later, Galileo wrote to answer a problem put to him by a gambler.

‡ M. G. Kendall points out to me that the problem posed by the Chevalier de Méré to Pascal concerning the problem of points involved similar small probabilities.

follow up * and from the way in which Fermat writes it seems clear that the actual definition of probability is assumed known. What the two savants were interested in was the application of this definition to specific problems which were concerned with dice playing between gamblers of equal skill and opportunity. The approach to the problems is similar to that of Galileo, and the generalizations which are made from the particular cases discussed are not well supported.

It is true that Galileo wrote on one problem only and fairly briefly at that, but it is difficult to see why Pascal and Fermat should be preferred as the originators of probability theory before Galileo or Cardano. It may well be that the precocity of Pascal as a mathematician led to much of his work being accepted with acclamation, and certainly without its priority being questioned. We find, for example, the famous Arithmetic Triangle in Stifel's *Arithmetica* (1543), in the *General Tratato* of Tartaglia in 1556, in the *Arithmetic of Simon Stevin of Bruges* (Leiden, 1625). It is possible that Pascal may not have known of these writers. However, he certainly knew of Pierre Herigone's *Cours Mathematique* (Paris, 1634), since he makes several references to it in his own *Usage du Triangle Arithmetique pour trouver les puissances des binômes et Apotômes*. Herigone uses a table of numbers analogous to the Arithmetic Triangle to find binomal coefficients. Perhaps this same aura which dazzled Pascal's contemporaries (and at the same time caused them to overlook some of Fermat's work) still blinds us to-day.

19. If we take the origins for granted and look at developments of the theory, then by far the greatest impetus to theory during the years 1650-60 must have come from the publication of *De Ratiociniis in Aleae Ludo* by Christian Huygens. Huygens as a young man of 26 arrived in Paris in July 1655 on the equivalent of the English 'Grand Tour'. He did not meet Pascal, Fermat or Carcavi, the intimate friend of Pascal, but he did meet Roberval, professor of mathematics at the Collège Royal de France, who is mentioned by Pascal as having been also approached by the Chevalier de Méré. Huygens stayed in Paris from July to November, and after his return to Holland he began a correspondence with both Carcavi and Fermat which lasted over a period of years. The young man's imagination was obviously fired by the discussions he had in Paris, and his mathematical ambitions stimulated by the immense activity of the group which some ten years later (1665) was to found the *Académie des Sciences*. He set himself to work, and in March 1656 he wrote to Prof. van Schooten that he had prepared a manuscript about dice games. Francis Schooten was professor of mathematics at Leyden and had been Huygens's teacher. He took the young Huygens's manuscript (which was written in his native language), translated it into Latin and published it as an appendix to his *Exercitationes Mathematicae* in 1657. (A French translation of this appendix can be found in *Oeuvres de Huygens*, tome 14, on 'Calcul des Probabilités' published by *La Société Hollandaise des Sciences* in 1920.) In this *Tractatus de Ratiociniis in Aleae Ludo* Huygens sets out in a systematic manner what he must have learnt in Paris and adds some results which he may have achieved himself.

In the letter to Francis Schooten he writes

...quelques-uns des plus Célèbres Mathématiciens de toute la France se sont occupés de ce genre de Calcul, afin que personne ne m'attribue l'honneur de la première Invention qui ne m'appartient

* It is interesting to see Pascal fall into the same kind of trap which caused D'Alembert such controversy. In discussing the game of heads and tails and the tossing of a coin D'Alembert argued that the probability of throwing a head with two tosses of a coin was 2/3. For we may have TT, TH or H—when we stop, the second throw being immaterial, since we have achieved what we want.

pas. Mais ces savants . . . ont cependant cachés leurs méthodes. J'ai donc dû examiner et approfondir moi-même toute cette matière à commencer par les éléments, et il m'est impossible pour la raison que je viens de mentionner d'affirmer que nous sommes partis d'un même premier principe. . . .

Accordingly Huygens begins by proving his basic propositions, deals at some length with the problem of points and then passes on to dice playing. His last proposition (XIV) has a familiar ring:

> If another gambler and I throw 2 dice turn and turn about with the condition that I will have won when I throw 7 points and he will have won when he throws 6, if I allow him to throw first, find my chance and his of winning.

His delineation of his fourteen propositions is admirably clear and concise, and it is no marvel that the tract was used by mathematicians as a reference book up to the time of James Bernoulli (who reprinted it) and beyond. Possibly by this crystallization of the ideas of the French mathematicians Huygens has earned the right to be regarded as the father of the probability theory.

20. After Huygens the interest of probabilists was not solely in gaming, although this interest did not die away for another hundred years or so. But with Huygens the new calculus seems fairly launched, and this is therefore a suitable point to make a break. There are many questions which one leaves unanswered. The drawings and paintings by palaeolithic man of himself are very rare, and there is probably no hope of finding pictures of his recreations. If he prized the astragalus as a toy it seemed a possibility that he might have carved or decorated it in some way, but I have not been able to find any record of this. But while we cannot pull aside the curtain from four hundred centuries the possibility does exist that the pre-historians may be able, one day, to take us back a little farther than the third millennium. The farther back one goes the more fragmentary the evidence, but the earliest dice found are described as being of 'well-fired buff pottery', and they certainly would not have been the first made.

The tantalizing period to the present writer is the period from the invention of printing to A.D. 1600. In this period we have two mathematicians only calculating probabilities, and yet this was in the immense intellectual ferment of the Italian Renaissance. It seems hardly possible that there were not other natural philosophers who attempted similar calculations, but such documents, if they exist, will only now come to light by chance.

The correspondence between the French mathematicians of the first half of the seventeenth century is almost complete, and presumably the possibility does exist here of finding further letters. They all seemed at one time or another to send letters to one friend under cover of letters to another, and such letters may conceivably still be ascribed to the wrong person. However, enough information does exist regarding the seventeenth-century mathematicians to make a coherent study, and if I appear to have done them scant justice it is because I find the period so interesting that I hope to write about it more fully on another occasion elsewhere.

Collecting information about dicing and gaming has been a hobby of mine for some time, and the list of persons who have drawn my attention to one aspect or another of it is formidable. I want to thank Prof. B. Ashmole of the British Museum who allowed me critically to examine the dice of the classical period which are in his care and M. Jean Charbonneaux of the Musée de Louvre who did me the same service. To Prof. C. M. Robert-

SOME FEATURES OF THE GENERATION TIMES OF INDIVIDUAL BACTERIA

By E. O. POWELL

Microbiological Research Department, Experimental Station, Porton

INTRODUCTION

The dynamics of bacterial growth in the mass has been the subject of a great deal of experimental study, but quantitative observations on the behaviour of individuals are scanty. In spite of the primary need for detailed knowledge of normal growth processes, it is the abnormal and morbid that have received most attention from both experimenters and theoreticians. Kendall (1952a) and Finney & Martin (1951) have recognized the necessity for extended studies of the normal condition; in this they have been prompted by the existence of two hypotheses (Kendall, 1948, 1952a; Rahn, 1932), both of which connect the observed scatter of generation times with an inner mechanism of fission.

Measurements of individual generation times sufficiently systematic and detailed to make analysis worth while have been provided only by Kelly & Rahn (1932) so far as I know. My object here is, first, to add to the existing corpus of data; secondly, to show that for technical reasons, the experimental work so far done (including my own) is not really adequate to test the hypotheses that have been proposed; thirdly, to suggest further lines of study.

According to Rahn's hypothesis, the fission of a bacterium involves the simultaneous duplication of a number of essential entities or structures which may be the genes. The time required for duplication will not be the same for each, but will be subject to a law of 'chance', and fission can only take place when every entity has been duplicated. Under certain assumptions, Rahn shows that the observed generation times should then be scattered according to the law

$$dF = \frac{g}{m} e^{-\tau/m} (1 - e^{-\tau/m})^{g-1} d\tau, \tag{1}$$

where dF is the proportion of generation times in the range τ to $\tau + d\tau$. The parameter g is equal to the number of entities which have to be duplicated, and m is a parameter depending on the rate of the individual processes. In what follows, I call the frequency function (1) 'Yule's distribution' (since Yule (1925), was the first to describe it).

Kendall's reasoning is similar, except that he supposes the events leading to fission to take place step by step, and fission to occur only when the series of steps is complete. Each step may or may not be a process of duplication. The observed generation times should then be scattered according to a Pearson Type III law in the form

$$dF = \frac{\tau^{g-1} e^{-\tau/m}}{m^g \Gamma(g)} d\tau,$$

where g is the number of steps and m is a kinetic parameter as before.

The idea of a distribution of generation times implies in practice (though this is not strictly necessary) some degree of permanence in the properties of the population to which it

applies. At a given instant, the organisms in a bacterial culture certainly possess a definite age distribution, but this has no significance if the growth rate and cultural conditions have varied during the life span of the oldest extant organism. *A fortiori*, a distribution of generation times, even if it be formally defined at an instant, will be a complex and imperspicuous character of its population unless the factors controlling it remain constant over at least several generations. Conditions adequate to the precise definition and determination of a distribution probably exist only, if at all, in a continuous culture maintained in a steady state. No measurements have yet been made on such a system, and I have assumed here (as has been done implicitly in the past) that during the phase of logarithmic growth in static culture the succession of generations proceeds in a regular manner. I give, however, two examples to the contrary.

It is usually found that the mass growth rate (the logarithmic derivate of the mass of organisms per unit volume) in a static culture rapidly accelerates at the end of the lag phase to a steady value which is maintained for several hours. The number growth rate (the logarithmic derivate of the number of organisms per unit volume) often behaves less regularly. It may be small or zero even after the mass growth rate has reached its maximum value; abnormally long organisms are thereby formed. There follows a stage of acceleration during which the number of growth rate approaches and temporarily exceeds the mass growth rate. The two then become equal, in some cases only for a short time before the phase of decline sets in. These remarks apply to culture in a liquid medium. Detailed quantitative information about the growth of colonies on solid media over long periods (i.e. of many generations) appears to be lacking, and it is only on solid media that observations of individual generation times can be made. With these difficulties in mind, it is clear that a satisfactory experimental procedure is not easily attained. Observations which are to be combined so as to furnish a distribution curve must either be confined to a brief period in the life of each culture examined, or, if extended, they must be shown to be drawn from a sensibly unchanging population.

In a culture whose number growth rate is varying there is temporarily an excess or deficit of young organisms; a change in the growth rate implies a change in generation time distribution. It does not necessarily imply a change in *all* the parameters of that distribution, but even if some one were known or assumed to be constant, it would be a matter of considerable practical difficulty to extract its value from a set of data derived from a changing population. At present it is not possible to test or to make use of Kendall's (1948) relation between the coefficients of variation of clone size and generation time; a method of maintaining a constant growth rate for many generations must first be established.

The results which I have to present consist of measurements of generation times carried out on six species of organisms (*Bacterium aerogenes*, strain N.C.T.C. 8197; *Bact. coli anaerogenes*, strain N.C.T.C. 4450; *Streptococcus faecalis*; *Proteus vulgaris*, strain LC; *Bacillus mycoides*, strain SR 2; *B. subtilis*, strain UP 1). Their distributions are compared with the frequency functions suggested by Kendall and Rahn. A partial analysis and discussion of some other features of the nexus of generation times are also given. I use the words 'mother', 'daughter', 'sister' with an obvious extension of their usual meaning, and the neutral terms 'inception', 'termination' to denote respectively the events at which an organism becomes a recognizably separate entity (by fission of its mother) and by which it ceases to be so (by itself dividing). The same terms may also refer to the epochs of these events. In describing the morphological changes in the cell wall occurring at fission I replace

Bisset's (1950) 'rough' and 'smooth' (which are overworked in bacteriology as well as in everyday use) by the mnemonic 'septate' and 'isthmoid' respectively (see de Bary, 1884; Schaudinn, 1902).

In view of the present uncertainty about the internal organization of the bacterial cell, it is misleading to speak of 'the nucleus' as a single discrete entity. I have retained the word for convenience, but it should be considered to mean no more, perhaps, than 'the complement of nuclear material'.

EXPERIMENTAL ARRANGEMENTS

(a) *Method of making observations: allowance for bias*

If observations of generation time are made on all the progeny of a single organism, and if the experiment is broken off at an arbitrary moment, the resulting sample will be biased in favour of short-lived organisms. For the organisms extant at the end of the experiment will have, in mean, a generation time greater than the average. The bias is by no means trivial, as I shall show. (Particular cases can be imagined in which these statements are not true; they are of no practical consequence.) An unbiased sample is obtained if observations are
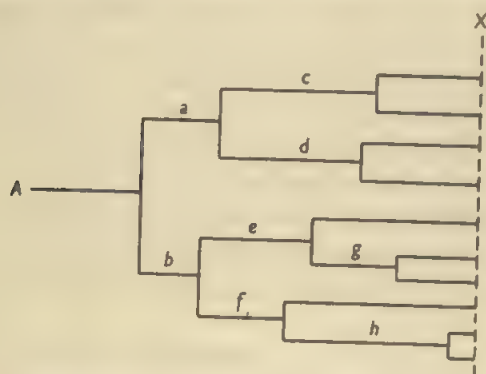
Fig. 1. A family tree of micro-organisms, illustrating bias.
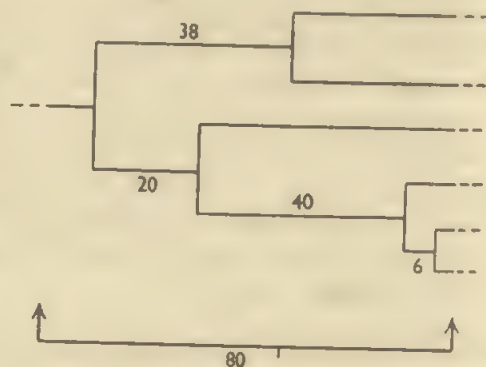
Fig. 2. A family tree of *B. mycoides*. The numbers give the life span of the organisms in minutes.

made of the same number of generations in every line of descent from the ancestor—provided that that number is decided beforehand and not allowed to be contingent upon the incidence of meal-times, for example. In the diagram of Fig. 1, the horizontal lines represent, against a uniform time scale, the life-spans of individuals in a family tree. The clone formed at the time X consists of six organisms of the third generation and four of the fourth; the times $a \ldots h$ form a biased sample, by inclusion of g and h. The times $a \ldots f$, i.e. two complete generations from the ancestor A, are without bias.

The method of observing every member of one or more trees of this sort, up to a fixed number of generations in every line of descent, was apparently adopted by Kelly & Rahn (1932), though their figures suggest that it was not adhered to quite rigidly. For organisms whose generation times are relatively little dispersed, the method is satisfactory.

When the scatter of generation times is large, the foregoing method becomes inefficient, for the total duration of a given number of generations is also widely scattered, and in the course of an experiment many possible observations will have to be rejected if bias is to be

avoided. An extreme example is shown in Fig. 2. To meet the difficulty, which arose in studying *B. mycoides* and *B. subtilis*, I adopted the device of systematizing experiments so that bias could be accepted and afterwards allowed for in a simple way.

Consider a group of organisms, n_0 in number at an arbitrary time $t = 0$, whose growth is followed for a period T (n_0 is supposed to be great enough for the growth to be regarded as continuous). In the period $0 \leqslant t \leqslant \xi$ there will be $n_0(e^{k\xi} - 1)$ fissions and so $n_0(e^{kT} - 1)$ in the whole period T. (Here k is the mean growth rate.) But an organism of generation time τ will terminate within the period T only if its inception occurs at a time $t < T - \tau$. Thus of all the organisms whose generation time is τ and whose inception lies within the period of observation, only a fraction

$$\alpha(\tau) = \frac{e^{k(T-\tau)} - 1}{e^{kT} - 1} \tag{2}$$

will on the average terminate, and so be measurable, in that period.

It is now possible to make use of all the observations available from studying the reproduction of organisms for a fixed time instead of for a definite number of generations. If the true distribution is

$$dF = f(\tau) \, d\tau,$$

the observed distribution will be

$$dF = B\alpha(\tau) f(\tau) \, d\tau, \tag{3}$$

with a necessarily finite range $0 \leqslant \tau \leqslant T$. The factor B is chosen so as to make

$$\int_0^T dF = 1.$$

Since generation times exceeding T are excluded, $f(\tau)$ can only be found as a distribution truncated at T; in practice the loss is negligible.

Accordingly, all the experiments on *B. mycoides* and *B. subtilis* were arranged to occupy a period of 80 min. (about three mean generations), so that the results could afterwards be combined and the above average allowance for bias made when fitting frequency functions.

The factor $\alpha(\tau)$ has a surprisingly large effect. For example, when τ has its mean value $\bar{\tau}$ and T is $3\bar{\tau}$, $\alpha(\tau)$ is $\frac{3}{7}$. With very large T, $\alpha(\tau)$ is nearly $\exp(-k\tau)$ and so is still as much as $\frac{1}{2}$ for $\tau = \bar{\tau}$. Its introduction is no matter of pedantry.

This reasoning can only be correct if (i) n_0 is very large; (ii) the culture has been growing steadily for a sufficiently long time for a stable age distribution to have been established; (iii) the n_0 are a fair sample from that distribution. Dr W. A. O'N. Waugh (private communication) has given a more exact treatment for cases in which these conditions are not met, but his calculations show that a few generations of growth are enough to satisfy (ii). The circumstances of the actual experiments and the tests applied for constancy of growth rate indicate that the error in taking $\alpha(\tau)$ as the biasing factor is negligible. It may very well be that there is always a human bias in the selection of organisms for observation, contrary to condition (iii), but it is scarcely possible to take account of it or to eliminate it without complicating the technique.

Observations of the growth of the selected organisms were made at intervals of 2 min., to correspond roughly with the uncertainty in estimating times of fission. The uncertainty in generation time was then about ± 3 min. This is an average figure only; the uncertainty was greater with organisms of unusually long generation time, which generally divided in a leisurely manner. On the other hand, unusually short generation times were estimated

much more accurately since inception and termination took place in close juxtaposition both temporally and spatially; records of generation times of 2, 4 and 6 min. may be taken as correct to ± 1 min., and therefore also correct as to their frequency. As it turns out, this is an important consideration.

The uncertainty introduces an error of a special kind. A generation time which should be recorded as τ may be recorded as $\tau - 2$ or $\tau + 2$ for instance: it will then appear in a frequency group adjacent to the correct one, and so there will be correlated errors in the frequencies. The effect will increase the variance, probably to a quite negligible extent, but there may also be a concomitant error in the χ^2 between the observations and a fitted function, making the fit seem better or worse than it is. I have found no way of estimating or allowing for this aberration.

I wish to make it quite clear what is here meant by a measured generation time τ: The fission of an organism is first observed to be complete, as nearly as can be judged, at a time t_1, say. It has therefore divided at some time between $t_1 - 2$ and t_1. Similarly, one of its daughters terminates at some time between $t_2 - 2$ and t_2. Then the generation time of that daughter is recorded as $t_2 - t_1$, which is always a multiple of 2. Thus τ is the mean value of the generation time of the group of organisms to which it is ascribed. I make this simple point because a careful study of Kelly & Rahn's (1932), Rahn's (1932) and Finney & Martin's (1951) papers raises doubts as to what Kelly & Rahn have actually recorded. Their figs. 1 and 2 suggest that they proceeded as I have done (though with a 5 min. instead of a 2 min. interval between observations), but in their tables the observations are grouped into ranges 0–5, 5–10, 10–15, etc. Rahn and Finney & Martin have the misleading phrase 'fissions ... observed in successive five-minute intervals'. Finney & Martin assumed that the mean generation times for these intervals were $2\frac{1}{2}$, $7\frac{1}{2}$, $12\frac{1}{2}$, ...; they may have been $2\frac{1}{2}$, 5, 10, 15, ... or 5, 10, 15, 20, ... I have verified if either of these alternatives is chosen, the goodness of fit of Yule's distribution is about the same as that obtained by Finney & Martin ($\chi_9^2 = 14\cdot9$), but the corresponding estimates of g are widely and significantly different: $18\cdot2$ and $36\cdot2$ respectively, as against $26\cdot0$.

Other observational arrangements are possible. For instance, the distribution of generation times can be deduced from the age distribution in a culture. If $\phi(a)$ is the frequency function of age and $f(\tau)$ that of generation time, it is found that if the culture is growing steadily

$$\phi(a) = \phi(0)\,e^{-ka} \int_a^\infty f(\tau)\,d\tau \tag{4}$$

(cf. Harris, 1951). Unfortunately, the form of the curve $\phi(a)$ is dominated by the value of k; it is very insensitive to changes in the parameters of $f(\tau)$ that leave k the same. Experimentally, the determination of an age distribution at a given time requires a study of the previous history of the culture over a period long enough to give the generation time distribution directly. I have, however, made use of the relation (4) in the reverse sense, as a check on the regularity of growth in experiments with *B. mycoides* and *B. subtilis*.

Again, suppose that a number of organisms extant at the epoch $t = 0$ are watched until they divide. The times $t = 0$ of fission will be distributed with a frequency function

$$\phi(0)\,e^{k\theta} \int_\theta^\infty e^{-k\tau} f(\tau)\,d\tau\,d\theta.$$

This gives another possible method of determining $f(\tau)$: I have made no use of it.

(b) *Practical details*

With one exception, organisms were grown on cellophane over tryptic meat broth in the culture chamber described by Harris & Powell (1951). The medium was constantly circulated so that an approximation to a condition of continuous culture obtained, the volume of medium (100 c.c.) was so large relative to the inoculum used (c. 100 organisms) that no appreciable change could occur during the course of an experiment, either by depletion of nutrients or accumulation of products of metabolism. In this respect the technique is superior to that of Orskov (1922) adopted by Kelly & Rahn. On the other hand, times of fission can be judged with less precision in cellophane culture; the average error appeared to be about ± 2 min. However, this contribution to the total variance is only a small fraction of that due to the organisms themselves.

The cultures were incubated throughout at a temperature of $35 \pm 0.5^\circ$ C. The two *Bacillus* species were inoculated as spores, the rest as saline dilutions from a 20 hr. liquid culture. To avoid confusion, the organisms on which observations were to be made were separated from their neighbours by means of a micromanipulator so that each lay within a clear area of cellophane. They were exposed to light as little as possible. Evaporation from the surface of the cellophane was controlled so that each organism or group of organisms was surrounded by a visible fillet of liquid (Harris & Powell, 1951). The evaporation was temporarily increased only when necessary to verify the occurrence of fission. A photographic method of recording was rejected after trial as less informative than direct examination, and no less tedious when the processing and assessment of films were taken into account.

No difficulty was encountered in working with *Bact. aerogenes* and *Bact. coli anaerogenes*. A typical experiment consisted in following the development of four organisms simultaneously for two or three complete generations. Organisms were selected for observation always at about 2 hr. from inoculation, previous experience having shown that erratic behaviour at the end of the lag phase would by then have subsided.

Strep. faecalis was found to be very sensitive to light; a few minutes exposure to the usual conditions of illumination sufficed to stop growth. The effect was minimized by adding 5 % of sheep serum to the medium and illuminating with red light as dim as could be tolerated. As a further precaution, observation was confined to one complete generation only of the progeny of the selected organisms. The numerical results suggest that these measures were effective in securing uniform unrestricted growth over the short period necessary. Observation was begun at $2\frac{1}{2}$ hr. from inoculation.

The mode of fission of *Pr. vulgaris* in cellophane culture was erratic, and seemed sometimes to be intermediate in character between the septate and the isthmoid. The first signs of a waist were apparently accompanied by the formation of a septum, and the organisms would remain for several minutes in this state of incipient fission. Attempts to assess generation times were unsatisfactory, and this species was therefore studied in culture on an agar-tryptic meat medium under incident dark-field illumination (Pearce & Powell, 1951). Judging by the lengths of organisms as then seen, the uncompleted fissions observed by the vertical illumination on cellophane were recorded as complete fissions under the dark field. For this organism, then, the conditions of culture were similar to those in Kelly & Rahn's experiments. For $3\frac{1}{2}$–4 hr. from inoculation it appeared to grow and reproduce steadily, but then the rate of fission declined, longer motile organisms were formed and swarming began. In order not to encroach on this phase of changing growth rate, observations were
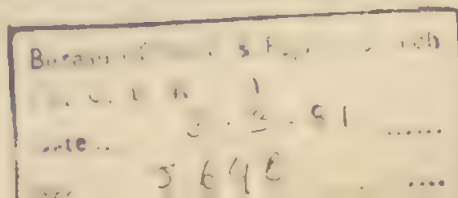
again confined to one generation only, beginning at $2\frac{1}{2}$ hr. from inoculation. (This organism did not swarm on cellophane over a circulating medium, presumably because no local concentration of metabolic products could be built up. See Lominski & Lendrum, 1947.)

Observations on *B. mycoides* were carried out during the 80 min. beginning at $2\frac{1}{4}$ hr. from inoculation. By that time the spore coat had been shed and families of two to four organisms developed. Three groups of experiments were conducted with *B. subtilis*: series (i), observation begun at $2\frac{1}{2}$ hr.; series (ii), observation begun at $4\frac{1}{2}$ hr.; series (iii), organisms growing over defibrinated sheep blood diluted with its own volume of normal saline; observation begun at 3 hr. In order to check the regularity of growth of these two species, the ages of the organisms extant at the end of each experiment were recorded, so that their distributions could be compared with those calculated by means of (4) from the generation time distributions.

THE FREQUENCY FUNCTIONS

I give in this section a statement of the frequency functions which are made use of below, together with their relevant properties. Each is limited to positive values of the variate, and each involves a parameter g which, whatever its interpretation, is a measure of intrinsic dispersion, i.e. a function of the coefficient of variation only.

Fitting was carried out either by the method of maximum likelihood or by that of moments (sometimes both). In several of the examples the distributions are such that the method of moments is of low efficiency (cf. Fisher, 1922). However, I do not think that the results are of much more than comparative value in any case, and the points I shall have to make are not so nice as to call for great statistical refinement.

(a) The Pearson Type III distribution (Kendall's hypothesis)

$$dF = \frac{\tau^{g-1} e^{-\tau/m}}{m^g \Gamma(g)} d\tau.$$

This is a well-known function. The cumulants are

$$\kappa_r = m^r g \Gamma(r),$$

whence the first three moments

$$\mu_1' = mg, \quad \mu_2 = m^2 g, \quad \mu_3 = 2m^3 g.$$

The coefficient of variation is

and the skewness and kurtosis

$$c_0 = g^{-\frac{1}{2}}$$

$$\gamma_1 = \kappa_3/\kappa_2^{\frac{3}{2}} = 2g^{-\frac{1}{2}}, \quad \gamma_2 = \kappa_4/\kappa_2^2 = 6/g.$$

Both tend to zero for large g.

Maximum-likelihood estimates $\hat{g}$, $\hat{m}$ of g and m are given by

$$\Sigma f\tau/n - \hat{g}\hat{m} = 0,$$
$$\Sigma f \log \tau/n - \log \hat{m} - \psi(\hat{g}) = 0,$$

where the f are observed frequencies of the generation times τ, n is Σf and $\psi(\hat{g})$ is

$$d \log \Gamma(\hat{g})/d\hat{g}.$$

(b) Yule's distribution (Rahn's hypothesis)

$$dF = \frac{g}{m}e^{-\tau/m}(1-e^{-\tau/m})^{g-1}\,d\tau.$$

Except for μ'_1, the moments are difficult to calculate and are best obtained through the cumulants (cf. Daniels, 1952; but see Yule, 1925). The cumulant generating function is easily found to be

$$K(t) = \log \Gamma(g+1) + \log \Gamma(1-itm) - \log \Gamma(g+1-itm),$$

whence

$$\kappa_r = (-1)^r m^r\{\Gamma_r(1) - \Gamma_r(g+1)\},$$

where for uniformity I write

$$\Gamma_r(x) = (d/dx)^r\{\log \Gamma(x)\}, \quad \Gamma_1(x) \equiv \psi(x).$$

In particular,

$$\mu'_1 = \kappa_1 = m\{\Gamma_1(g+1) - \Gamma_1(1)\}, \tag{5}$$

$$\mu_2 = \kappa_2 = m^2\{\Gamma_2(1) - \Gamma_2(g+1)\}, \tag{6}$$

$$c_0 = \frac{\{\Gamma_2(1) - \Gamma_2(g+1)\}^{\frac{1}{2}}}{\Gamma_1(g+1) - \Gamma_1(1)},$$

$$\gamma_1 = \frac{\Gamma_3(g+1) - \Gamma_3(1)}{\{\Gamma_2(1) - \Gamma_2(g+1)\}^{\frac{3}{2}}}.$$

The curve is distinguished from a Type III having the same first two moments by its slower rise near zero, and by its greater steepness on the left flank. In fact, for increasing g, γ_1 diminishes but tends to a still rather large non-zero limit

$$-\pi^{-3}6^{\frac{1}{2}}\,\Gamma_3(1) = 1\cdot14.$$

In practice γ_2 scarcely differs from $2\cdot4\ (= 3!(\pi^4/90)\,(6,\pi^2)^2)$.

Some other features are discussed by Finney & Martin (1951) and Rahn (1932), who note the insensitivity of the form of the curve to changes in g.

The maximum-likelihood equations for fitting are

$$\frac{n}{\hat{g}} + \Sigma f\log(1 - e^{-\tau/\hat{m}}) = 0,$$

$$-\frac{n}{\hat{m}} + \Sigma f\tau/\hat{m}^2 - \frac{\hat{g}-1}{\hat{m}}\Sigma\frac{f\tau/\hat{m}}{e^{\tau/\hat{m}} - 1} = 0.$$

The solution is greatly facilitated by the tables of Einstein functions due to Sherman & Ewell (1942).

(c) The Pearson Type III distribution with allowance for bias

From equation (3), if the true distribution is of Pearson Type III, that derived from data which are biased in the way already described will be

$$dF = B\alpha(\tau)\frac{\tau^{g-1}e^{-\tau/m}}{m^g\Gamma(g)}\,d\tau, \tag{7}$$

where $\alpha(T)$ is given by equation (2). This is a frequency function in the range $0 \leqslant \tau \leqslant T$ ($T = 80$ in all experiments); the moments strictly involve incomplete gamma functions, but

these are here so nearly unity that no appreciable error arises in taking the range of integration as infinite. And then it is found that

$$B = \left\{\frac{e^{kT}(1+km)^{-g}-1}{e^{kT}-1}\right\}^{-1},$$

$$\mu_1' = mg\left\{\frac{e^{kT}(1+km)^{-g-1}-1}{e^{kT}(1+km)^{-g}-1}\right\},$$

$$\mu_2' = m^2g(g+1)\left\{\frac{e^{kT}(1+km)^{-g-2}-1}{e^{kT}(1+km)^{-g}-1}\right\},$$

with $k = (\log 2)\,(mg)$. The parameters are hence obtained by iteration; approximate values of m and g are inserted into the expressions in curly brackets, and the two equations for the moments solved to give a closer approximation which is used as the starting point for a second cycle, and so on. For the *terminus a quo*, it is enough to take $m = 1$, and $\bar{\tau} = gm = 25\,\text{min}$. (a value roughly known *a priori*).

(d) Yule's distribution with allowance for bias

$$dF = B\alpha(\tau)\frac{g}{m}e^{-\tau/m}(1-e^{-\tau/m})^{g-1}\,d\tau. \tag{8}$$

With the same approximation as in the previous paragraph,

$B = (e^{kT}-1)/H,$

$$\mu_1' = m\{\Gamma_1(g+1)-\Gamma_1(1)\}\left[H\cdot\frac{\Gamma_1(1+km+g)-\Gamma_1(1+km)}{\Gamma_1(g+1)-\Gamma_1(1)}-1\right]\Big/(H-1),$$

$$\mu_2' = m^2[\Gamma_2(1)-\Gamma_2(g+1)+\{\Gamma_1(g+1)-\Gamma_1(1)\}^2]$$

$$\times\left[H\frac{\Gamma_2(1+km)-\Gamma_2(1+km+g)+\{\Gamma_1(1+km)-\Gamma_1(1+km+g)\}^2}{\Gamma_2(1)-\Gamma_2(g+1)+\{\Gamma_1(g+1)-\Gamma_1(1)\}^2}-1\right]$$

$$\times[H-1]^{-1},$$

where
$$k = (\log 2)/m\{\Gamma_1(g+1)-\Gamma(1)\},$$

$$H = e^{kT}\frac{\Gamma(g+1)\,\Gamma(1+km)}{\Gamma(g+1+km)}.$$

After fitting the function (7) to a set of data, estimates of the true moments were calculated from its parameters, and were used to obtain values of the parameters of (8) by means of (5) and (6). These values were taken as the first approximation in fitting (8), iteration being carried out as in the previous section.

ANALYSIS OF THE DATA

General

The fission of a micro-organism, as usually understood, consists in its separation into two geometrically distinct pieces, each with a closed cell wall separating it from its surroundings. The closure of the wall follows the division of the cytoplasm at some unknown interval, but in unicellular organisms there is at least a unique one-to-one correspondence between the true cell division and the observed fission. In the genus *Bacillus* this is not so; the organisms are multicellular, and at any time contain party walls of various degrees of completeness.

The party walls may or may not be laid down as double layers, but they cannot be seen to be so, and it is not known that each becomes necessarily a site of fission, in these organisms fission is a tertiary rather than a secondary process (for recent work on the structure of the party wall, see Dawson & Stern, 1954). Thus the generation times of *B. mycoides* and *B. subtilis* have not the significance for the hypotheses of Rahn and Kendall that those of the other organisms have. (The same applies to Kelly & Rahn's measurements on *B. cereus*. Rahn recognized this.) There is no *a priori* reason why either hypothesis should not apply, as regards its mathematical form, to the *Bacilli*, but the 'genetic' parameter *g* cannot then be interpreted as the number of genes or primitive steps determining fission. For the moment, little more than a descriptive treatment of the generation times of these organisms can be usefully attempted in this context.

The raw data are set out in Table 1, together with the first two moments and the skewness and kurtosis of each distribution. The range of generation times is very great, and much at variance with the vague assumption often made, that bacteria divide rather regularly. A striking feature in all the experiments was the highly erratic behaviour of small families of organisms (e.g. the six or fourteen constituting the progeny, up to two or three generations, of a single ancestor); in some the generation times are quite uniform, in others widely dispersed. This unforeseen lack of homogeneity makes statistical analysis difficult, and renders doubtful the significance of any broad and simple treatment. Such quantitative analysis as I have been able to apply can only be a makeshift; but a more refined investigation will not be justified until a larger range of data has accumulated.

In the paragraphs which follow, some forward references are unavoidable; homogeneity is considered first because it affects the methods adopted in assessing the parameter *g*, but the results of curve-fitting are also relevant, and have been used to modify the tests of homogeneity.

Homogeneity of the data

Because of the special difficulties associated with *B. mycoides* and *B. subtilis*, only the four unicellular organisms *Bact. aerogenes*, *Bact. coli anaerogenes*, *Strep. faecalis* and *Pr. vulgaris* will be considered in this section.

Kendall (1948) by applying Bartlett's (1937) test for homogeneity of variance, showed that the estimates of his *g* obtained from the several experiments of Kelly & Rahn were highly inconsistent; and later (1952 *a*) he expressed doubts as to the propriety of combining the data to obtain an overall estimate of *g*. Such doubts are of course justified if the source of variation is a change in the experimental conditions, and certainly Kelly & Rahn's results strongly suggest that day-to-day fluctuations are considerable.

Any one experiment of the present series yields measurements of the generation times of organisms belonging to several families, each the progeny of a single organism selected at the beginning of the experiment. The organisms all lie within a single field of the microscope and are subjected therefore to nearly identical environmental conditions. The variance of generation times within the experiment must be that proper to the organism under those conditions, and the coefficient of variation (a constant for each species by hypothesis) should yield an unbiased measure of *g*.

Now it remains true of the present results, as well as of Kelly & Rahn's, that there is apparent lack of homogeneity between experiments. An examination of the coefficients of variation showed them to be fairly uniformly distributed, except for a few which were outstandingly large; the principal contribution to the variance in these cases was provided by

Table 1. *Frequency of generation times, with the first (μ_1', $=\tau$) and second (μ_2) moments and the measures γ_1, γ_2 of skewness and kurtosis*

Generation times	Bact. aerogenes	Bact. coli anaerogenes	Strep. faecalis	Pr. vulgaris	B. mycoides	B. subtilis series (i)	B. subtilis series (ii)	B. subtilis series (iii)
0	—	—	—	—	—	—	—	—
2	—	—	—	—	—	—	—	—
4	—	—	—	—	2	4	1	1
6	2	3	1	—	4	6	2	2
8	3	8	0	4	6	8	2	3
10	4	7	0	4	10	15	6	2
12	13	24	1	2	12	13	3	3
14	35	45	4	5	19	26	13	6
16	48	49	6	9	13	24	14	15
18	75	56	13	7	20	32	23	27
20	72	47	8	20	25	34	24	35
22	65	53	15	34	22	42	25	25
24	40	39	9	42	27	27	25	40
26	42	30	20	42	23	39	34	44
28	29	17	12	41	17	19	30	32
30	9	16	11	29	12	22	32	28
32	13	8	8	34	9	13	28	28
34	2	3	4	26	14	15	19	18
36	4	4	3	25	5	17	24	14
38	1	7	0	17	6	8	6	12
40	1	0	0	19	11	1	11	12
42	2	4	1	13	9	1	11	11
44	0	1	0	1	4	3	7	9
46	2	—	0	4	3	—	3	1
48	—	—	1	3	4	—	1	5
50	—	—	0	1	1	—	1	1
52	—	—	1	0	1	—	2	3
54	—	—	—	0	0	—	2	0
56	—	—	—	2	0	—	1	1
58	—	—	—	2	1	—	2	1
60	—	—	—	3	0	—	1	0
62	—	—	—	—	0	—	—	2
64	—	—	—	—	0	—	—	—
66	—	—	—	—	1	—	—	—
68	—	—	—	—	0	—	—	—
70	—	—	—	—	0	—	—	—
72	—	—	—	—	0	—	—	—
74	—	—	—	—	0	—	—	—
76	—	—	—	—	1	—	—	—
78	—	—	—	—	—	—	—	—
80	—	—	—	—	—	—	—	—
Σf	462	421*	118	390	282	369	353	381
No. of expts.	14	12	9	19	16	19	18	16
No. of families measured	51	51	59	93	49	43	53	66
μ_1'	21·05	20·48	24·90	27·25	22·90	20·38	25·72	25·64
μ_2	32·94	43·98	46·37	75·43	119·76	66·40	89·64	83·99
γ_1	0·79	0·72	0·67	0·57	0·87	0·18	0·45	0·66
γ_2	1·78	0·84	2·07	1·25	1·60	−0·46	0·70	1·12

* One organism failed to grow and suffered lysis.

one or (rarely) more families of very unequal generation times. Thus for example with *Bact. coli anaerogenes* in an experiment involving four families each of six organisms, the mean generation times were

$$26\cdot3, \quad 21\cdot7, \quad 28\cdot3, \quad 27\cdot0;$$

the variances

$$51\cdot8, \quad 2\cdot2, \quad 56\cdot6, \quad 135\cdot5;$$

and the squared coefficients of variation

$$0\cdot0748, \quad 0\cdot00467, \quad 0\cdot0707, \quad 0\cdot186.$$

Irregularities of this kind, which were found in all the species examined, appear to make it doubtful that any simple meaning can be attached to the coefficient of variation of larger samples.

In order to test the homogeneity of within-family variance, the Bartlett criterion, often denoted by M (which for normal samples is distributed approximately as a χ^2), was computed for each experiment in the usual way and the sum, ΣM, formed for each group of experiments. A correction for kurtosis was applied, given by Box (1953) as

$$\Sigma M' = \Sigma \left\{ M \Big/ \left(1 + \frac{1}{2} \frac{\nu}{\nu+1} \gamma_2 \right) \right\}.$$

This expression was simplified by taking a common corrective factor $(1 + \frac{1}{2}\gamma_2)$ for every family, so biasing M' towards smaller values.

Two assumptions were made in assigning γ_2: (i) If the underlying distribution is of Type III, γ_2 is $6/g$. In this case the γ_2 were calculated from the final estimates of Kendall's g given by curve fitting (Table 9). (ii) If the underlying distribution is Yule's, γ_2 is very nearly $2\cdot4$. The results of the test given in Table 2 show that the apparent heterogeneity is not significant under Rahn's hypothesis, but is markedly so under Kendall's. To this extent it might be argued that Rahn's hypothesis is to be preferred. But examination of the figures in detail shows that the large contributions to M' come mainly from families containing members of unusually short generation time, whereas it is the long upper tail of Yule's distribution which inflates the fourth moment; near the origin the ordinates are exceedingly small. There is, however, no sufficient reason here to suspect Kelly & Rahn's experiments of being ill-controlled.

In order to assess the extent to which the overall variance of generation time is increased by inconstant experimental conditions, I have carried out conventional analyses of variance on the means of families. From Table 3 it can be seen that in every case a significant part of the total variance is contributed by differences in the growth rate from experiment to experiment. If σ_F^2 and σ_E^2 are the true within and between experiment variances, the observed between-experiment mean square will be an estimate of

$$\sigma_F^2 + \frac{N^2 - \Sigma n^2}{\bar{N}(K-1)} \sigma_E^2$$

(where K is the number of experiments, N the number of families, and n the number of families in an experiment). On this basis we derive the figures in the last column of Table 3 as estimates of the experimental variance of the growth rate. Except for *Pr. vulgaris*, these estimates of σ_E^2 are not great; nevertheless, estimates of the coefficient of variation from the raw data will be too high.

Table 2. Modified Bartlett test of homogeneity of variance of families

	Bact. aerogenes	Bact. coli anaerogenes	Pr. vulgaris
(A) Kendall hypothesis: γ_2	0.38	0.55	0.43
$\Sigma M'$	75	78	109
n	28	39	72
$P(\Sigma M')$	0.000	0.000	0.000
(B) Rahn hypothesis: γ_2	2.4	2.4	2.4
$\Sigma M'$	41	45	60
n	28	39	72
$P(\Sigma M')$	> 0.05	> 0.1	> 0.1

Table 3. Analysis of variance of family means

Organism	Source of variation	Sum of squares	Degrees of freedom	Mean square	F-ratio (and $P(F)$)	Estimated σ_E^2
Bact. aerogenes	Between expts.	305·95	13	23·53	2·03 (0·05)	3·30
	Within expts.	429·32	37	11·60		
	Total	735·27	50			
Bact. coli anaerogenes	Between expts.	314·16	11	28·56	2·86 (<0·01)	4·38
	Within expts.	389·34	39	9·98		
	Total	703·50	50			
Strep. faecalis	Between expts.	610·18	8	76·27	2·37 (<0·05)	6·86
	Within expts.	1609·21	50	32·18		
	Total	2219·39	58			
Pr. vulgaris	Between expts.	2037·45	18	113·19	3·58 (<0·0001)	17·23
	Within expts.	2274·42	72	31·59		
	Total	4311·87	90			

By contrast, the coefficients of variation of generation times of families, with the exception of *Pr. vulgaris*, are not significantly partitioned. The coefficients themselves (and their squares) are very unsymmetrically distributed, but the distribution of their logarithms, as shown by a graphical test, is roughly normal, and analyses of variance carried out on the logarithms yields the results of Table 4. For convenience, the variable was taken as $100\log_{10} c^2$ to the nearest integer; the constant factors of course cancel in the *F* ratio. *Strep. faecalis* again had to be omitted from this test, because many families, each consisting of a pair of sisters only, had zero variance. However, the test was also carried out on the squared coefficients of variation themselves; the sense of the results was exactly the same, and *Strep. faecalis* gave an *F* ratio probability above 0·1

Table 4. *Analysis of variance of family coefficients of variation*

Organism	Source of variance	Sum of squares	Degrees of freedom	Mean square	F ratio and P/P.
Bact. aerogenes	Between expts.	20386	13	1568	0·84 (-)
	Within expts.	67246	36	1868	
	Total	87632	49		
Bact. coli anaerogenes	Between expts.	17668	11	1606	0·86 (-)
	Within expts.	72701	39	1864	
	Total	90369	50		
Pr. vulgaris	Between expts.	159425	18	8857	2·94 (0·001)
	Within expts.	216582	72	3008	
	Total	376007	90		

Thus the coefficient of variation is relatively stable as between experiments, but its large interfamily variance is difficult to account for on Kendall's hypothesis if, as is supposed, the generation time of an organism is independent of its immediate ancestry. It may well be a product of 'delayed fission'; that is, the lapse of an appreciable interval between the division of nucleus or cytoplasm, and the separation of cell wall. A single large delay in a family postpones the termination of one organism and the inception of two others, and so may have a disproportionate effect on the coefficient of variation. To admit this possibility is to entertain the view that the measurements may not be directly relevant to either hypothesis.

The experiments with *Pr. vulgaris* were evidently less successful than the rest as regards reproducibility of growth rate; this is perhaps to be associated with the organisms' potentially more complex behaviour on the static medium used.

Correlation between generation times

Kelly & Rahn (1932) suggested that some of the more extreme values occurring among their measurements were to be accounted for by 'delayed division'. Such delay would obviously tend to generate negative correlation between mother and daughter and positive correlation between sisters. Rahn (1932) showed that this tendency existed, but he did not consider the measurements as a whole. It is probable that delayed division in this sense is too rigid a conception, and that a delay of variable magnitude always occurs; this does not affect the conclusion.

The product-moment correlation coefficients between generation times of mothers and daughters (ρ_{MD}) and pairs of sisters (ρ_{SS}) in the present series of measurements are given in Table 5. Not all the mother-daughter pairs are independent, since in a family of three or four generations some organisms appear twice as mothers and once as daughters. Scatter diagrams of the mother-daughter correlations showed the entries to be fairly uniformly

Table 5. *Correlation between generation times of mothers and daughters (ρ_{MD}) and sisters (ρ_{SS}). Number of pairs of observations, p. Parentheses indicate that the coefficient is not significantly different from zero at the 5 % level*

Organism	ρ_{MD}	p	ρ_{MD} corr. for bias	ρ_{SS}	p	ρ_{SS} corr. for bias
Bact. aerogenes	(0·069)	360	—	0·77	231	—
Bact. coli anaerogenes	(−0·051)	318	—	0·48	210	—
Strep. faecalis	—	—	—	0·68	59	—
Pr. vulgaris	—	—	—	0·74	195	—
B. mycoides	−0·18	192	−0·42	0·42	103	0·32
B. subtilis, series (i)	(0·090)	283	(0·016)	0·76	156	0·80
B. subtilis, series (ii)	−0·37	252	−0·34	0·65	154	0·59
B. subtilis, series (iii)	(−0·005)	248	−0·16	0·60	169	0·36

clustered about their centroid. This was true also for the bulk of sister-sister pairs; there was, however, a tail to the distribution extending along the bisector of the angle between the axes, indicating very strong association between pairs of sisters of unusually long generation time.

For *Bact. aerogenes* and *Bact. coli anaerogenes* ρ_{MD} is not significantly different from zero, and the immediate conclusion is that delay does not play a large part in determining the variance of generation time. If this is true, neither does it contribute much to the correlation between pairs of sisters (ρ_{SS}) which is yet quite large. There is no constraint implicit in the hypotheses of Kendall and Rahn which requires sister cells to be alike; in fact, independence is assumed. The high value of ρ_{SS} tells heavily either against the hypotheses, or against the relevance to them of the observations. Now ρ_{MD} is compounded of observations made on a number of different occasions. It is to be expected that the true mean value of τ is slightly different for each experiment, and so if ρ_{MD} were really zero a small positive value might be obtained from the combined data. It is possible therefore that the conditions bias both

ρ_{SS} and ρ_{MD} in favour of positive values. ρ_{MD} may be appreciably negative, and ρ_{SS} smaller than it appears to be. However, inspection of the records leaves no doubt as to the great similarity between sister organisms.

In the *Bacillus* group, ρ_{MD} is either zero or negative, but ρ_{SS} remains positive to a high degree of significance. The septate mode of fission appears to be a more highly organized process than the isthmoid mode, and it probably gives rise to very appreciable delay. Its contribution to ρ_{MD} cannot be assessed without knowledge of its variance, though selected examples suggest that the variance may be extremely large, especially in *B. mycoides*. But the differences between the three series of *B. subtilis* indicate that ρ_{MD} is far from being a simple property of the organism. The growth rate in series (ii) and (iii), as will be shown below, was not constant; this should introduce a positive bias, yet ρ_{MD} for (ii) is the most negative of all.
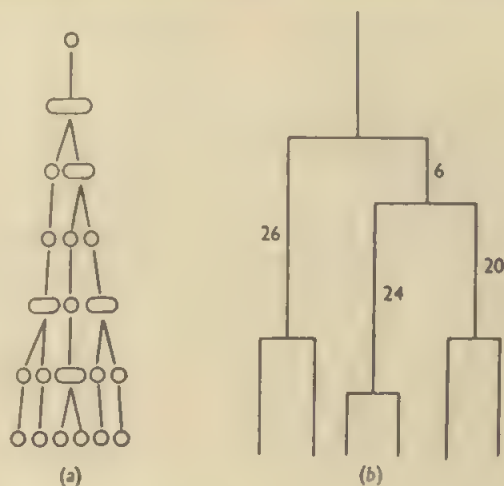
Fig. 3.   Abnormal development of *Strep. faecalis*.   (a) Successive appearance of cells.
(b) Corresponding family tree; generation times in minutes.

The estimates of ρ_{SS} and ρ_{MD} are affected by the presence of bias in the frequencies. I have attempted a correction in the following way.

(i) If M and D are the generation times of mother and daughter, and $R(M, D)$ the observed frequency of the two together, an estimate of the true frequency is

$$R(M, D)/\alpha(M + D),$$

where α is the biasing factor of equation (2).

(ii) If S_1, S_2 are the generation times of a pair of sisters, and $R(S_1, S_2)$ the frequency of the pair, the corrected value is $R(S_1, S_2)/\alpha\{\max(S_1, S_2)\}$. The ρ calculated from frequencies corrected in this way are also shown in Table 5. No new regularity appears, and the sense of the foregoing remarks is not altered.

Once delayed fission is admitted as a possible contribution to the total variance of generation time, it becomes a matter of importance to assess its variance, since it may turn out so large as to render nugatory any attempt on the present lines to test the hypotheses of Kendall and Rahn. A single overt example of delay was found among the observations on *Strep. faecalis*. Fig. 3a shows the appearances; a single cell, instead of dividing when it had reached the usual size, continued to grow until its length was about three times its diameter.

It then split into a sphere and short rod, and fission of the rod followed in 6 min. Fig. 3*b* is the family tree, with the times of fission placed so as to correspond horizontally with Fig. 3*a*. This example is sufficient to engender suspicion that minor degrees of delay may be much more frequent.

Some further progress might be made by a more detailed analysis of correlation among the generation times. I have not attempted such analysis because a direct experimental approach should be possible through the use of ultra-violet or phase-contrast illumination.

I think it likely that the observed values of ρ_{MD} will be found to result from two opposing effects: a small positive correlation, corresponding to the usual succession of τ values, over-laid by less frequent successions of great disparity. Taken together with the large dispersion of within-family variance, this suggests the following picture: a very large family tree, representing the numerical and temporal extension of a culture, consists of a patchwork of areas within some of which growth is steady, while within others it is erratic. There is a kind of correlation between mothers and daughters, but it is not such as the crude product-moment coefficient is well suited to describe.

The generation time distributions

(a) B. mycoides *and* B. subtilis

Although no quantitative conclusions relevant to the hypotheses of Kendall and Rahn can be drawn from the data on these species, the observations have an important bearing on the theory of fission. The results of fitting the modified Type III and Yule distributions

Table 6. *Frequency function parameters and goodness of fit*

Frequency function		B. mycoides	B. subtilis series (i)	B. subtilis series (ii)	B. subtilis series (iii)
Pearson Type III with allowance for bias (eq. (7)):	g	4·07	6·11	7·01	7·47
	m	7·04	3·81	4·21	3·90
	χ^2	19·2	57·3	17·8	15·4
	n	17	15	16	16
	$P(\chi^2)$	0·32	0·000	0·34	0·50
Yule, with allowance for bias (eq. (8)):	g	4·64	9·48	—	—
	m	13·3	8·18	—	—
	χ^2	22·2	74·2	—	—
	n	18	14	—	—
	$P(\chi^2)$	0·22	0·000	—	—

(equations (7) and (8)) are given in Table 6 (see also Fig. 6). Series (i) of *B. subtilis* relates to observations begun at $2\frac{1}{2}$ hr. from inoculation, series (ii) at $4\frac{1}{2}$ hr., series (iii) to observations on organisms growing over diluted sheep blood.

Not only do these species exhibit a higher dispersion of generation time than the rest; the lengths of the organisms are also correspondingly variable, though the parallel between

generation time and length at inception or termination is not exact. The first few vegetative organisms to grow from the spore are often very long—upwards of 30μ—and although the mean length later diminishes, the change is not so well marked as in say *Bact. aerogenes* at the beginning of the logarithmic growth phase. *B. mycoides* cannot be said to possess a normal mean length; the organisms become progressively shorter over a period which may be as long as 24 hr. in some media. While a diminution in mean length is occurring the number growth rate must be greater than the mass growth rate, and since it is initially less, there is acceleration at some stage. I have therefore checked the uniformity of growth during measurements of generation time in the following way:

It will be recalled that all the experiments grouped into any one column of Table 1 were begun at the same epoch in the life of the culture, and each occupied 80 min. Number growth
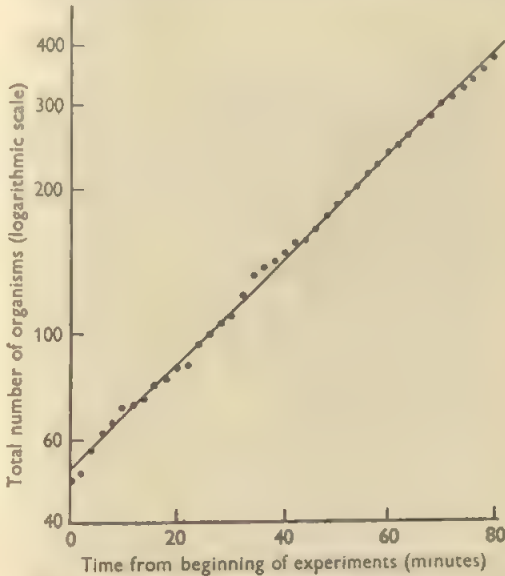
Fig. 4. Number growth rate of *B. mycoides*.
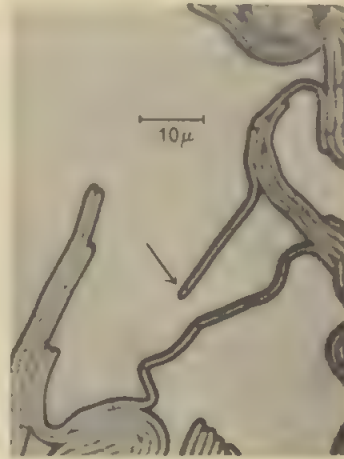
Fig. 5. *B. mycoides* growing on cellophane, 5 hr. from inoculation. A single cell has split off from one end of a long organism.

curves were constructed by adding together the number of organisms under observation at corresponding times in each experiment. For *B. mycoides* and *B. subtilis* series (i) the logarithm of this sum was a linear function of time (Fig. 4), but for *B. subtilis* series (ii) and (iii), there was marked curvature especially over the earlier part of the period. Numerical evaluation of the age distribution by means of equation (4) was also carried out, and compared with the observed age distribution at the 80 min. epoch; series (ii) and (iii) showed an obvious excess of young organisms, as would be expected from an increasing growth rate. These two series are therefore of no quantitative value.

The extraordinarily wide range of generation times in *B. mycoides* (Table 1) is a concomitant of its readiness to divide at any available point (i.e. between any pair of cells). The exaggerated appearance of Fig. 5 is by no means a rarity; the frequency with which a single cell is split off from an organism of eight or more cells is diagnostic of the species. Such single-celled organisms are viable, but develop rather slowly, and some do not divide within 3 hr. of their inception; they are not spores (cf. Bergersen, 1954). *B. subtilis* shows

3

something of the same freedom in its earlier stages of growth, but it reverts more quickly to regularity. Clearly, in these species the generation time is intimately connected with the coarser structure of the organism, and only remotely with the nuclear fission.

The distribution of *B. subtilis* at 2½ 4 hr. (series (i)) is notable for its approximate symmetry (Table 1), in which respect it is quite outstanding; neither the Type III nor Rahn's distribution gives an acceptable fit.

The experiment of growing *B. subtilis* over blood was an attempt to test the constancy of the coefficient of variation of generation time. Pearce and I had found (1951) that in these circumstances the organisms were very short and uniform in size, which suggested that the dispersion would also be small. The blood used in the present work was diluted with saline in order to reduce its viscosity, and over this mixture long organisms were first formed as usual. The mean length gradually diminished, but at the intermediate stages both long and
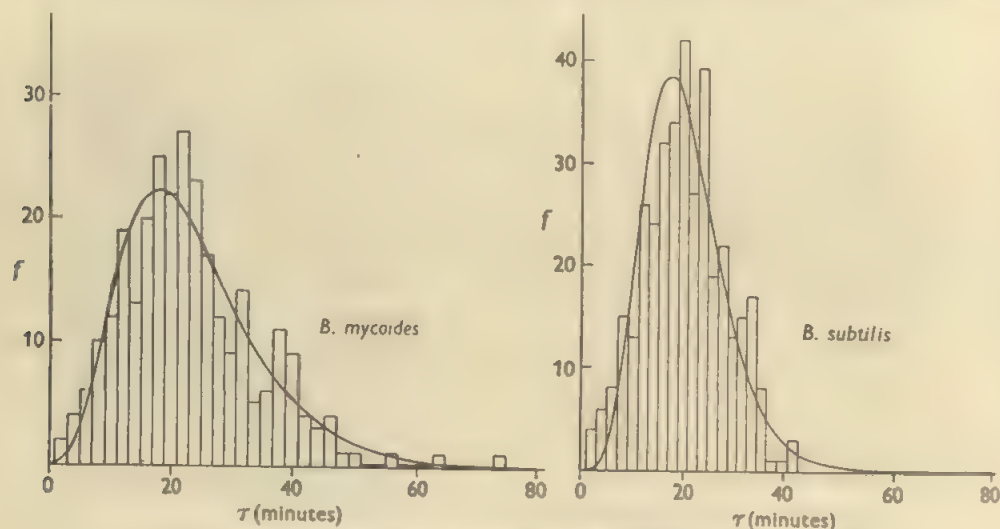
Fig. 6. Generation time distributions of *B. mycoides* and *B. subtilis*, with modified Pearson Type III frequency functions.

short organisms were present, each clone being predominantly of one sort or the other. As I have shown, the growth rate increased considerably during the course of observation. Despite the imperfection of the experimental conditions, the coefficient of variation (0·36) was rather less than in the other two series (0·40 and 0·37).

These rather disconnected remarks indicate that the behaviour of *B. mycoides* and *B. subtilis* is so complex that the experiments are quite insufficient to expose it fully. The distributions, however, possess one regular feature worth examination.

Fig. 6 shows that the frequency of small generation times in *B. mycoides* and *B. subtilis* series (i) is greater than the expectation accorded by the assumed frequency function (a modified Type III). This is true also for *B. subtilis* series (ii) and (iii) (and has since been observed in *B. megatherium*). The observed frequencies near the origin are uniformly so large as to suggest that the true frequency function approaches zero with a non-zero slope (I have pointed out that the observations are less liable to error in this region than elsewhere). In no case did two successive fissions occur in the same 2 min. interval. It will be seen from Fig. 6 that no Pearson Type III or Yule function could give a tolerable fit over the

higher range without being in defect at low values of τ. Both of these functions have a finite positive and non-zero slope at the origin only when $g = 2$ and then

$$\left.\frac{df}{d\tau}\right|_{0} = \frac{1}{m^2} \quad \text{and} \quad \frac{2}{m^2} \quad \text{respectively.}$$

But the distributions with $g = 2$ have values of γ_1 and γ_2 quite different from those observed. Kelly & Rahn's results for *B. cereus* are equally suggestive, their grouped frequencies, in order of increasing mean τ, are 9, 2, 4, 5, 17, ...

Among the unicellular organisms, one instance (*Strep. faecalis*, $\tau = 6$ min.) has already been noted of a generation time which was unusually short because of delayed fission in the previous generation. It is perhaps significant that this strain, like the *Bacilli*, exhibited the septate mode of fission. But it cannot be said of *B. mycoides* that the shortest generation times are associated with anomalous development, for reasons already given. Many instances of short generation time in *B. subtilis* can, however, be accounted for as resulting from an unusual succession of fissions. Fig. 7a represents the usual succession; an organism about to divide has typically three completed party walls (represented by pairs of dots). At a time t_0 it divides at or near its centre, and the two daughters similarly at t_{11} and t_{12}. For simplicity, the organism is shown as dividing without growing. The intervals $t_{11} - t_0$ and
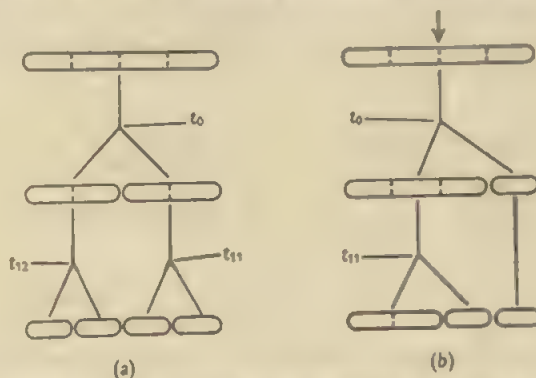
Fig. 7. Normal (a) and abnormal (b) succession of fissions in *B. subtilis*.

$t_{12} - t_0$ are not generally extreme. Occasionally (Fig. 7b) the first division occurs at a distance of about one-quarter of the length of the organism from one end, and a second division follows very shortly; then the interval $t_{11} - t_0$ is small. That half of the organism to the right of the arrow in Fig. 7b can be usefully regarded as an organism *manqué* with a negative generation time $t_0 - t_{11}$, i.e. the associated separation of cell walls has occurred in the opposite order to the corresponding separation of cytoplasm.

I believe that the impression given by the generation time distributions of *B. mycoides* and *B. subtilis*, namely, that the slope does not vanish near the origin, is a just one, and that the phenomenon is of fundamental significance for the mechanism of fission in *Bacillus*; I hope to deal with it more fully in another communication.

(b) Bact. aerogenes, Bact. coli anaerogenes, Strep. faecalis, Pr. vulgaris; *Assessment of the genetic parameter*

The result of fitting the Type III and Yule distributions to the observations on unicellular organisms is shown in Table 7 A–D.

Comparison of sections A and C shows at once that the Type III distribution fits the data better than does Yule's, in each separate case. Moreover, the method of moments gives (section B) substantially the same results as the method of maximum likelihood. Applied to Yule's distribution, it gives a rather better fit, with very different values of the parameters (section D). This disagreement, with the moment fit better than the maximum likelihood, seems to indicate that Yule's distribution is inappropriate to the raw data.

However, the analysis of variance has shown that the crude coefficient of variation (c) of τ must exceed the supposed true and constant value (c_0) because of the experimental dispersion of the mean, and so the method of moments (Table 7 B and D) underestimates g —sampling fluctuations apart. As a further consequence, it does not follow from these results that Kendall's hypothesis is to be preferred to Rahn's.

Kelly & Rahn's results on *Bact. aerogenes* have been shown by Finney & Martin (1951) to be in satisfactory agreement with Rahn's hypothesis so far as the χ^2 test can distinguish. I have already pointed out that there is some doubt as to the correct assignment of mean generation time to the grouped frequencies, but with the same assignment as that adopted by Finney & Martin, viz.

$$\tau = 2\tfrac{1}{2} \quad 7\tfrac{1}{2} \quad 12\tfrac{1}{2} \quad 17\tfrac{1}{2} \quad \dots$$
$$f = 0 \quad 1 \quad 7 \quad 61 \quad \dots$$

I find that the Type III curve fits less well than Rahn's and shows no advantage in its estimate of c and γ_1 (Table 8). With other assignments of generation times to frequencies

$$\tau = 2\tfrac{1}{2} \quad 5 \quad 10 \quad 15 \quad \dots$$
$$f = 0 \quad 1 \quad 7 \quad 61 \quad \dots$$

and

$$\tau = 5 \quad 10 \quad 15 \quad 20 \quad \dots$$
$$f = 0 \quad 1 \quad 7 \quad 61 \quad \dots$$

the estimated parameters are different, but the sense of the results is the same.

Hypothesis requires that the genetic parameter (and therefore the true coefficient of variation of τ) shall remain constant despite variations in growth rate. Finney & Martin (1951) pointed out that the best estimate of it would be obtained by fitting frequency functions to the data for each experiment separately, constraining g to be the same for each, but allowing m to differ as required from experiment to experiment. As they realized, this would be an exceedingly laborious process with Rahn's frequency function, but it turns out that for the Type III function a simple maximum-likelihood solution exists.

It is convenient to write the Type III function in the form

$$dF = \frac{g^g \tau^{g-1} e^{-g\tau/a}}{a^g \Gamma(g)} \, d\tau,$$

where the new parameter a is the mean of the distribution, equal to gm. Let a_r be the value of a to be assigned for the rth experiment. Then the likelihood for that experiment is

$$\log L_r = (g-1)\Sigma(f\log\tau) - g\Sigma f\tau/a_r - g(\log a_r)\Sigma f + g(\log g)\Sigma f - \{\log \Gamma(g)\}\Sigma f,$$

where the f are the observed frequencies of τ in the rth experiment.

On summing over all experiments and differentiating we obtain the desired maximum-likelihood estimate of g:

$$n\{\psi(\hat{g}) - \log(\hat{g})\} = \sum_r \{\Sigma(f\log\tau)\} - \sum_r \{(\log\hat{a}_r)\Sigma f\}, \tag{9}$$

Table 7. *Frequency function parameters and goodness of fit*

Frequency function and method of fitting		*Bact. aerogenes*	*Bact. coli anaerogenes*	*Strep. faecalis*	*Pr. vulgaris*
(A) Pearson Type III, maximum likelihood:	g	13·4	9·48	12·8	9·08
	m	1·57	2·16	1·95	3·00
	χ^2	16·1	15·1	11·6	27·1
	n	11	12	9	15
	$P(\chi^2)$	0·14	0·23	0·24	0·028
(B) Pearson Type III, moments:	g	13·4	9·53	13·4	9·84
	m	1·57	2·15	1·86	2·77
	χ^2	16·1	15·3	10·2	23·4
	n	11	12	9	15
	$P(\chi^2)$	0·14	0·22	0·33	0·076
(C) Yule, maximum likelihood:	g	32·9	17·1	26·1	14·2
	m	5·21	5·99	6·56	8·48
	χ^2	29·4	24·1	20·0	44·9
	n	12	13	9	16
	$P(\chi^2)$	0·0033	0·030	0·018	0·0002
(D) Yule, moments:	g	59·9	27·2	59·2	29·6
	m	4·50	5·23	5·34	6·84
	χ^2	22·0	13·2	15·9	42·5
	n	10	11	8	14
	$P(\chi^2)$	0·015	0·28	0·043	0·0001
(E) Pearson Type VI, first two moments and σ_E^2:	g	15·1	10·7	15·9	13·1
	s	136	97·7	92·4	45·1
	b	2848	1980	2274	1202
	χ^2	15·4	15·8	11·3	30·1
	n	10	11	8	14
	$P(\chi^2)$	0·12	0·15	0·19	0·0074
(F) Yule-hyperbolic (eq. (14)), first two moments and σ_E^2:	g	79	36	92	55
	l_1	24·3	24·3	29·7	35·1
	l_2	18·1	17·1	20·6	20·6
	χ^2	21·2	14·7	15·1	40·7
	n	10	10	7	13
	$P(\chi^2)$	0·020	0·14	0·035	0·0001

where
$$\hat{a}_r = \Sigma f\tau / \Sigma f,$$

and n is $\sum\limits_r (\Sigma f)$, the total number of observations. The mixed derivatives

$$\frac{\partial^2}{\partial g \, \partial a_r} \quad \text{and} \quad \frac{\partial^2}{\partial a_r \, \partial a_s} \quad (r \neq s)$$

of $\sum\limits_r \log L_r$ are all zero (this is the virtue of using the parameter a instead of m), and so, as in the simple case, the asymptotic variance of $\hat{g}$ can be estimated as

$$\operatorname{var} \hat{g} = 1/n\{\Gamma_2(\hat{g}) - 1/\hat{g}\}.$$

Table 8. *Comparison of Pearson Type III and Yule distributions applied to Kelly & Rahn's results for Bact. aerogenes*

	g	m	χ^2	n	$P(\chi^2)$	c	γ_1
Type III	10·8	2·90	17·0	8	0·030	0·30	0·61
Yule	26·0	8·14	14·9	9	0·094	0·33	1·18
From data	—	—	—	—	—	0·31	0·91

The results of applying equation (9) to the data for the unicellular organisms are shown in Table 9 E. The g values obviously differ significantly from species to species, and all are less than 20, the approximate figure arrived at by Kendall (1948) from a consideration of Kelly & Rahn's experiments on *Bact. aerogenes*. In spite of the fact that the variance of τ in the present measurements on that organism is less than in theirs, the value of g is still only 16·5; Kendall, of course, was not aware that the apparent heterogeneity of the data was largely intrinsic.

Although the preceding form of treatment is not available for Rahn's hypothesis, a satisfactory approach can in any case be made through the coefficient of variation.

A number of generation time experiments carried out on one species can be considered to provide a sample of frequency functions $h(\tau, a, c_0)$ drawn from a population in which the true coefficient of variation $c_0 = \sigma/a$ is by hypothesis constant, but the mean a is dispersed. In this hypothetical population of h-functions, let $j(a)$ be the frequency function of the mean. Then considering the whole series of experiments together, the expectation of an observation τ will be

$$dF = \int_{a_l}^{a_u} h(\tau, a, c_0) \, j(a) \, da \, d\tau, \tag{10}$$

where a_u and a_l are the limits within which a must lie. The moments about zero are then

$$\mu_r' = \int_0^\infty \int_{a_l}^{a_u} \tau^r h(\tau, a, c_0) \, j(a) \, da \, d\tau.$$

It can safely be assumed that the order of integration may be inverted, so that

$$\mu_r' = \int_{a_l}^{a_u} \int_0^\infty j(a) \, \tau^r h(\tau, a, c_0) \, d\tau \, da.$$

In particular,

$$\mu_1' = \int_{a_i}^{a_u} a j(a)\, da = \mathbf{a}, \quad \text{say.}$$

$$\mu_2' = \int_{a_i}^{a_u} j(a)\, a^2 (c_0^2 + 1)\, da = (c_0^2 + 1)\,\mu_2'(j). \quad \text{say.}$$

Hence the observed crude coefficient of variation, c, is related to c_0 by

$$c^2 = \mu_2'/(\mu_1')^2 - 1 = (c_0^2 + 1)\,\mu_2'(j)/\mathbf{a}^2 - 1.$$

(Evidently $c \geqslant c_0$ always, since $\mu_2'(j) \geqslant \mathbf{a}^2$.) Or, more symmetrically, writing c_j for the coefficient of variation of the j-distribution,

$$c^2 + 1 = (c_0^2 + 1)(c_j^2 + 1). \tag{11}$$

Thus from a crude coefficient c a corrected value c_0 can be calculated if the experimental variance of the mean is known, always under the hypothesis that c_0 is constant. The necessary figures are already available from the working of the analysis of variance; the σ_E^2 of Table 3 are estimates of $\mu_2(j)$ and

$$c_j^2 = \mu_2(j)/\mathbf{a}^2.$$

Application of equation (11) then yields the figures in Table 9 B.

It is to be noted that the relation between g and c_0 so calculated is determined by the h-distribution, and is independent of $j(a)$.

If from the corrected values c_0 new estimates of Kendall's $g\,(= 1.c_0^2)$ are calculated, the results agree well with the revised maximum likelihood estimates (Table 9 C and E), and so the assumption that the h-distribution is of Type III is not an unreasonable one.

It is now possible to derive also improved values of Rahn's g from

$$c_0^2 = \frac{\Gamma_2(1) - \Gamma_2(g+1)}{\{\Gamma_1(g+1) - \Gamma_1(1)\}^2}.$$

These values are given in Table 9 D.

Equation (10) suggests a method of developing modified frequency distributions capable of representing the crude data more accurately, and of being subjected to a meaningful significance test. Each series of experiments furnishes a number of values a_r and a variance σ_E^2. The exact distribution of a is not known, but a graphical examination of the individual a_r shows a marked modal tendency. Suppose then, for convenience, that a is distributed as a Pearson Type V variate:

$$j(a) = \frac{b^s a^{-s-1} e^{-b/a}}{\Gamma(s)},$$

with

$$\text{overall } \bar{\tau} = \mathbf{a} = b/(s-1),$$

and

$$\sigma_E^2 = \mu_2(j) = b^2/(s-1)^2 (s-2).$$

Then if the distribution $h(\tau, a, c_0)$ is of Type III, equation (10) is easily integrated, and becomes

$$dF = \frac{(b/g)^s}{B(g,s)} \frac{\tau^{g-1}}{(\tau + b/g)^{g+s}}\, d\tau, \tag{12}$$

where $B(g, s)$ is the complete beta-function. This is a Pearson Type VI distribution.

I have fitted equation (12) to the raw data by using the first two moments and the variance σ_E^2 of the mean, to determine b, s and g (Table 7 E; the values of g are of course the

same as those in Table 9 C. since the j-distribution is irrelevant). To neglect σ_E^2 and use the first three moments would omit part of the information available, viz. the knowledge of how the τ were grouped by experiments. The goodness of fit is no better than that of the simple Type III function. One of the fitted curves—*Pr. vulgaris*, for which σ_E^2 is largest— is shown in Fig. 8 (curve (i)).

Table 9. *The crude (c) and corrected (c_0) coefficients of variation,*
with final estimates of the genetic parameter (g)

	Bact. aerogenes	Bact. coli anaerogenes	Strep. faecalis	Pr. vulgaris
(A) c	0·273	0·324	0·273	0·319
(B) c_0	0·257	0·306	0·251	0·277
(C) g (Kendall) from c_0	15·1	10·7	15·9	13·1
(D) g (Rahn) from c_0	79	36	92	55
(E) g (Kendall) by maximum likelihood	16·5 ± 1·1	10·9 ± 0·7	18·6 ± 2·3	14·4 ± 1·0

Another convenient choice of $j(a)$ is to take it as rectangular with mean a and range $2\sqrt{3}\,\sigma_E$. Then, still with a Type III distribution for $h(\tau, a, c_0)$, equation (10) becomes

$$dF = \frac{g}{2\sqrt{3}\,\sigma_E(g-1)}\left\{I\left(\frac{\tau g}{a+\sqrt{3}\,\sigma_E}, g-1\right) - I\left(\frac{\tau g}{a-\sqrt{3}\,\sigma_E}, g-1\right)\right\}d\tau, \qquad (13)$$

where $I(x, p)$ is the incomplete gamma function

$$\frac{1}{\Gamma(p)}\int_x^\infty \xi^{p-1}e^{-\xi}\,d\xi.$$

This distribution, fitted to the *Pr. vulgaris* data, is also shown in Fig. 8 (ii). The χ^2 is 29·5, as against 30·1 for the Type VI, and the close geometrical similarity of the two curves shows that it matters little what functional form is chosen for $j(a)$. *A fortiori* this is true for the other organisms, for which σ_E^2 is much smaller.

Yule's distribution does not combine readily, in equation (10), with any of the standard distributions. However, after the result of the previous paragraph, we may take

$$j(a) = \frac{\lambda}{a} \quad (l_1 \geqslant a \geqslant l_2),$$

i.e. $j(a)$ is a segment of a rectangular hyperbola. Then

$$\lambda = 1/\log(l_1/l_2),$$
$$a = \lambda(l_1 - l_2),$$
$$\sigma_E^2 + a^2 = \mu_2'(j) = \lambda(l_1^2 - l_2^2).$$

Equation (10) becomes

$$dF = \int_{l_2}^{l_1}\frac{\lambda g^\Psi}{a^2}e^{-\tau\Psi/a}(1 - e^{-\tau\Psi/a})^{g-1}\,da\,d\tau,$$

where $\Psi = \Gamma_1(g+1) - \Gamma_1(1)$ by equation (5), $a = m^2\Psi$, and by hypothesis Ψ is constant. Integration is immediate on writing $x = e^{-\tau\Psi/t}$

$$dF = \frac{\lambda}{\tau}[\{1-e^{-\tau\Psi/t_1}\}^g-\{1-e^{-\tau\Psi/t_1}\}^g]d\tau. \qquad (14)$$

Like the Type VI, this distribution fits no better than the parent Yule distribution with smaller g (Table 7 D and F).
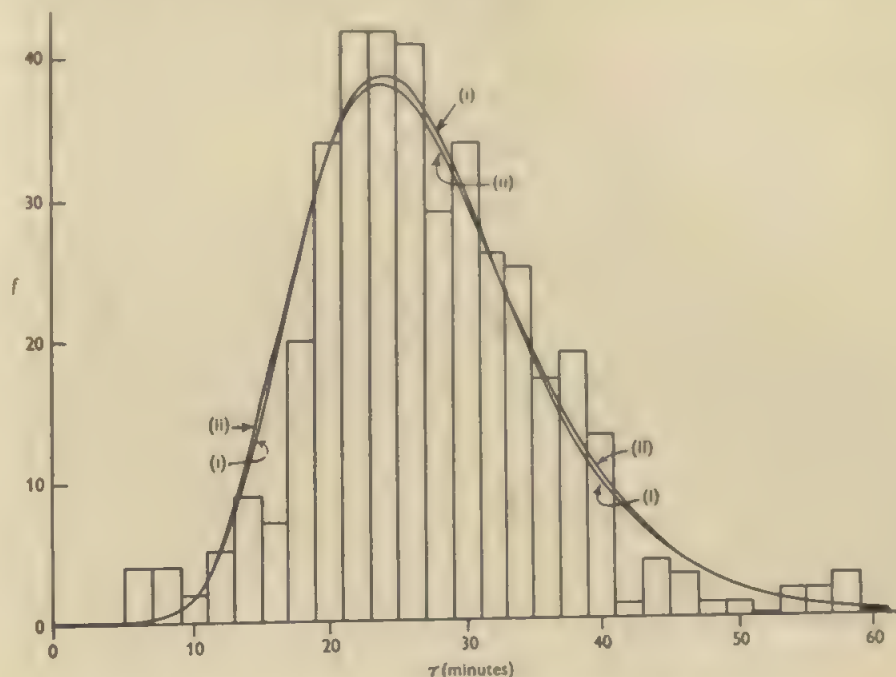
Fig. 8. Generation time distribution of *Pr. vulgaris* with fitted curves.
(i) Pearson Type VI, (ii) distribution of equation (13).

As in the *Bacillus* group, there is in all four species an excess of observed over expected frequencies near the origin (Table 10); this is true for all the distributions of Table 7. The organisms of very short generation time are well distributed among the several experiments carried out on each species; thus the ten *Pr. vulgaris* organisms of $\tau \leqslant 10$ (see Fig. 8) are distributed among five experiments—they are not associated with a single experiment of exceptionally low mean τ. And it is evident that if the restriction, $c_0 = $ constant, is maintained, only extravagant dispersion of the mean could appreciably increase the ordinates of the fitted curves near the origin. Qualitatively, this is just such an effect as might result from the occurrence of occasional large delays in fission; because of the skewness of the distribution the presence of a short-lived daughter organism would be much more obvious than that of its long-lived mother.

On the score of goodness of fit, therefore, Kendall's hypothesis is to be preferred to Rahn's, but the proposed frequency functions do not represent the observations altogether satisfactorily. The large scatter of within-family variance (Table 2) is admissible only on Rahn's

hypothesis, but it is associated with the presence of an excess of organisms of very short generation time, and cannot be solely accounted for by the long upper tail of the Yule distribution.

Table 10. *Observed and expected frequencies of short generation times*

(The range of τ is, for uniformity, taken to be half the distance of the mode from the origin.)

	Bact. aerogenes	Bact. coli anaerogenes	Strep. faecalis	Pr. vulgaris
Range of τ	0–9	0–9	0–11	0–11
Observed frequency	5	11	1	10
Expected, Type III fitted by moments	1·7	6·8	0·6	3·4
Expected, Yule fitted by moments	0·04	1·6	0·02	0·4
Expected, Type VI	1·4	5·9	0·5	2·2
Expected, Yule-hyperbolic (eq. (14))	0·03	1·2	0·02	0·3

CONCLUSIONS

The values of the corrected coefficients of variation in Table 9, and the corresponding estimates of the 'genetic' parameter g constitute the principal quantitative outcome of this study. Other features of the pattern of generation times, though less precisely expressible, are not less important.

Kendall's g is nowhere as much as 20, and obviously cannot be identified with the number of genes in the organism. Kendall himself (1952a) does not in fact insist on any particular interpretation. Rahn's g ranges up to about 100, which is much larger than the estimate of Finney & Martin (25 for *Bact. aerogenes*); however, private opinion among geneticists appears to indicate that this number is still absurdly low, and their view certainly seems justified by the wide range of genetically determined properties already known in *Bact. coli* and *Neurospora crassa* for example.

The observations on *B. mycoides* and *B. subtilis*, though possessing a specialized interest of their own, are not at present susceptible of any simple interpretation, however tentative.

It is very likely that a mechanism of Kendall's type, i.e. a stepwise process, does occur during the fission of an organism, and there is no *a priori* reason why it should not also be preceded or accompanied by Rahn's gene-duplication process. Then the one will be manifested clearly only if it is slow enough, relative to the other, to be the dominant factor in determining generation time. If neither is dominant (and this is a real possibility) the experimental picture will be confused, though it may not be recognized to be so. As things stand, Kendall's hypothesis is to be preferred, but only on the ground that its formal expression as a Pearson Type III distribution is in fair accord with the data. Kendall and Waugh (Kendall, 1952a) have also proposed a modification of Kendall's original hypothesis by relaxing the condition that the duration of the primitive steps of his fission process shall have the same frequency function for all. The modified hypothesis subsumes Rahn's as a special case (though only formally, not structurally). I have not examined its practical consequences in detail, but it has one significant feature pointed out by Kendall: suppose g is estimated by fitting a Type III distribution to a set of data, then if the primitive steps are not all distributed alike, their number will be greater than g.

It would obviously be possible to test Rahn's hypothesis by measuring generation times of the same species at different temperatures and when growing on different media: the coefficient of variation should remain constant. Under Kendall's more reserved hypothesis, change of temperature would be expected to make no difference, but his g might well depend on the number of synthetic processes demanded by the laying down of nuclear matter or cell wall, for instance, and so might depend in turn on the chemical complexity of the nutrients offered. Further, a similar study of a wide range of organisms, including auto-trophs and the most exigent pathogens, would show whether or not there is any correlation between the dispersion of generation time and richness of capacity and structure. Rahn (1932), reaching to the mammals for an extreme comparison, suggests that there is such a correlation.

But I do not think that studies of this kind would be at all profitable at present. There are serious objections not only to the acceptance of either hypothesis, but also to the acceptance of the data as critical of either: defects in the frequency functions, correlation between generation times of sister cells, heterogeneity of variance. The hypotheses, at least in their primary intention, relate to nuclear processes whose effect is seen only at two or three removes. At every point of difficulty in the foregoing discussions the possibility naturally suggests itself that delayed division plays an appreciable part in the dispersion of generation time; that the recognizable termination of the cell succeeds the essential determinative process by an interval which is itself sensibly dispersed. Therefore, the immediate need is for improvement in technique. The use of ultraviolet illumination should enable nuclear fission to be seen directly, and so permit the generation time so-called to be analysed into its components. Extended study of cultures on cellophane over a flowing medium would also be valuable: when constant growth rate can be reliably maintained over many generations, it will be possible to calculate the dispersion of generation time simply from the dispersion of clone size, with great economy in time and patience (Kendall, 1948). Further mathematical work is also in progress (see Kendall, 1952b).

SUMMARY

1. The conditions are discussed which must be met in any attempt to measure a frequency distribution of generation times of micro-organisms.

2. Generation times of individual organisms of six species have been measured: *Bacterium aerogenes*, *Bact. coli anaerogenes*, *Streptococcus faecalis*, *Proteus vulgaris*, *Bacillus subtilis*, *B. mycoides*.

3. The hypotheses developed by Kendall and by Rahn each connect dispersion of generation time with a postulated mechanism of fission; each implies a definite mathematical form for the distribution. Comparison with experiment suggests that Kendall's is to be preferred.

4. In Rahn's hypothesis, one of the parameters of the distribution is identified with the number of genes in the organism. When every allowance is made for experimental error, and in the most favourable case (*Strep. faecalis*), the estimated number of genes is less than 100.

5. Generation times are not distributed at random, but are in part determined by a weak and variable hereditary factor effective only over a few generations.

6. There are positive objections to the acceptance of either Kendall's or Rahn's hypo-thesis: correlation between generation times of sister cells, heterogeneity of variance in

small families of organisms, certain defects in the proposed distribution functions. These difficulties, and the effect mentioned under (5), may be due to delayed fission, that is, to the lapse of an appreciable period between the division of the nucleus and the observed separation of the cell into two parts.

7. In the *Bacillus* group, the measurements are not relevant to the hypotheses, because the organisms are multicellular.

8. Improved techniques are required for the further pursuit of the subject.

REFERENCES

BARTLETT, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc.* A, **160**, 268–82.

DE BARY, H. A. (1884). *Vergleichende Morphologie und Biologie der Pilze, Mycetozoen, und Bakterien.* Leipzig.

BERGERSEN, F. J. (1954). The filamentous forms of *Bacillus megaterium. J. Gen. Microbiol.* **11**, 175–9.

BISSET, K. A. (1950). *The Cytology and Life-history of Bacteria.* Edinburgh: E. and S. Livingstone.

Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, **40**, 318–35.

DANIELS, M. E. (1952). The covering circle of a sample from a circular normal distribution. *Biometrika*, **39**, 137–43.

DAWSON, I. M. & STERN, H. (1954). Changes in the bacterial cell wall during cell division as seen in the electron microscope. *J. Gen. Microbiol.* **10**, iii.

FINNEY, D. J. & MARTIN, L. (1951). A re-examination of Rahn's data on the number of genes in bacteria. *Biometrics*, **7**, 133–44.

FISHER, R. A. (1922). On the mathematical foundation of theoretical statistics. *Philos. Trans.* A, **222**, 309–68.

HARRIS, NORA K. & POWELL, E. O. (1951). A culture chamber for the microscopical study of living bacteria, with some observations on the spore-bearing aerobes. *J. R. Micr. Soc.* **71**, 407–20.

HARRIS, T. E. (1951). Some mathematical models for branching processes. *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability*, p. 305.

KELLY, C. D. & RAHN, O. (1932). The growth rate of individual bacterial cells. *J. Bact.* **23**, 147–53.

KENDALL, D. G. (1948). On the role of a variable generation time in the development of a stochastic birth process. *Biometrika*, **35**, 316–30.

KENDALL, D. G. (1952a). On the choice of a mathematical model to represent normal bacterial growth. *J. R. Statist. Soc.* B, **14**, 41–4.

KENDALL, D. G. (1952b). Les processus stochastiques de croissance en biologie. *Ann. Inst. Poincaré*, **13**, 43–108.

LOMINSKI, I. & LENDRUM, A. C. (1947). The mechanism of swarming of *Proteus. J. Path. Bact.* **59**, 688–91.

ORSKOV, J. (1922). Method for the isolation of bacteria in pure culture from single cells and procedure for the direct tracing of bacterial growth on a solid medium. *J. Bact.* **7**, 537–49.

PEARCE, T. W. & POWELL, E. O. (1951). New techniques for the study of growing micro-organisms. *J. Gen. Microbiol.* **5**, 91–103.

RAHN, O. (1932). A chemical explanation of the variability of the growth rate. *J. Gen. Physiol.* **15**, 257–77.

SCHAUDINN, F. (1902). Beiträge zur Kenntnis der Bakterien und verwandter Organismen. I. *Bacillus bütschlii*, n.sp. *Arch. Protistenk.* **1**, 306–43.

SHERMAN, J. & EWELL, R. B. (1942). A six-place table of Einstein functions. *J. Phys. Chem.* **46**, 641–61.

YULE, G. U. (1925). A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S. *Philos. Trans.* B, **213**, 21–87.

QUANTUM HYPOTHESES

By S. R. BROADBENT

The British Coal Utilisation Research Association

1. INTRODUCTION

(1·1) Statistical problems which involve a mixed population are complex, even when the component subpopulations can be assumed to be normal. Karl Pearson (1894) noted in his dissection of a frequency curve into two normal components, 'It may happen that we have a mixture of $2, 3, ..., n$ homogeneous groups, each of which deviates about its own mean symmetrically, and in a manner represented with sufficient accuracy by the normal curve.... The equations for the dissection of a frequency curve into n normal curves can be written down as for the special case of $n = 2$ treated in this paper, they require us only to calculate higher moments. But the analytical difficulties, even for the case of $n = 2$, are so formidable, that it may be questioned whether the general theory could ever be applied in practice to any numerical case.'

The present paper discusses the simpler problem in which the means of the components are equally spaced. The hypothesis that this is the case has been called by Hammersley & Morton (1954) a *quantum hypothesis*, for the means are then a constant plus multiples of a basic quantity or quantum.

Examples of such distributions will be found in the paper and in the references cited. It is worth noting that such rules as Brook's law (1886), formulated by Fowler (1909) in his study of the Ostracoda, that 'during early growth. each stage increases at each moult by a fixed percentage of its length, which is approximately constant for its species and sex', and Przibram's rule (1912) discussed by Wigglesworth (1942), that 'the weight doubles at each instar, and at each moult all linear dimensions are multiplied by $\sqrt[3]{2}$', take the form of quantum hypotheses when the logarithms of the weights or lengths are taken. Therefore situations in which a quantum hypothesis applies after the variate is transformed are here considered.

(1·2) Data suggest a quantum hypothesis by the occurrence of regularly spaced modes. From such data three results are commonly required:

(i) an estimate of the quantum which determines the spacing of the modes,

(ii) an estimate of the scatter within the subdistributions, and

(iii) a demonstration that the quantum rule is not disobeyed, i.e. that the data are not more likely to come from some other distribution, perhaps unimodal.

When the valleys between the modes are very well marked, there is little difficulty in meeting these requirements. Controversy about the interpretation of some data shows that a statistical treatment is required for those cases in which grouping at the modes is less obvious.

In the most difficult situation we are presented with data alleged to support a quantum hypothesis. No independent information is available; the hypothesis has arisen from a study of the data and the means are estimated from the observations. We are required to test whether the observations are genuinely grouped about these means. For examples of this type, see Hammersley & Morton's (1954) Druid Circle problem and Grant's (1952) measure-

ments of the energy levels of atomic nuclei. Hammersley & Morton conjecture that in this form the problem is beyond present-day analytic resolution, and offer a Monte Carlo method of testing the supposed grouping. In this case the difficulty is that the data which prompted the hypothesis are also used for estimation and for testing, and it is not clear what allowance can be made in a test for such previous use. The warning given by Pearson & Chandrasekar (1936) is relevant: 'To base the choice of the test of a statistical hypothesis upon an inspection of the observations is a dangerous practice; a study of the configuration of the sample is almost certain to reveal some feature, or features, which are exceptional if the hypothesis is true....By choosing the feature most unfavourable to the hypothesis out of a very large number of features examined, it will usually be possible to find some reason for rejecting the hypothesis.' We need add only that it will be equally possible to choose some untrue hypothesis and to find some test which favours it.
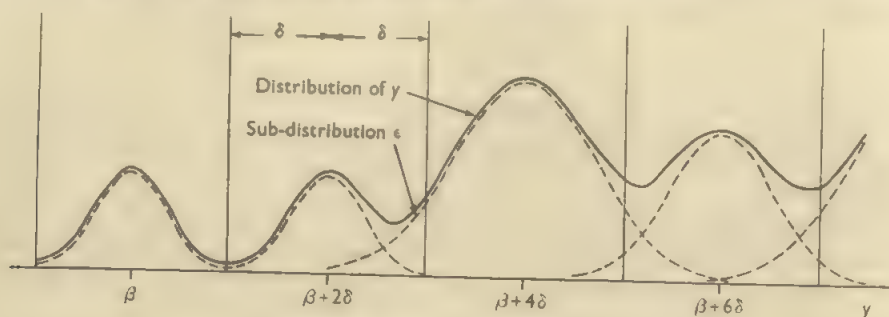
Fig. 1. Frequency curve on the quantum hypothesis.

(1·3) The population we shall consider will be compounded of normal subdistributions with means equally spaced at $\beta + 2r\delta$ ($r = 0, 1, ..., m$). The probability that a randomly chosen member of the population is from the tth subdistribution, i.e. has expectation $\beta + 2t\delta$, is p_t; not all the p_t are supposed non-zero, but $\sum_{t=0}^{m} p_t = 1$. The frequency function of such a population is shown in Fig. 1. The subdistributions may be homoscedastic, with common variance σ^2, or the s.d. may increase linearly with the mean.

(1·4) The frequency histogram of n observations from such a population will usually consist of regularly spaced peaks if σ/δ is small and the p_t are not too different. As σ/δ increases the valleys between the peaks disappear and it becomes difficult to assign an observation to its correct subdistribution. In these circumstances the p_t determine the overall appearance of the population, i.e. whether it is unimodal, symmetrical and so on. It is clear that using this model and by suitable choice of the parameters we can approximate with any required precision to any distribution, and that the graduation between multimodal and unimodal distributions is continuous.

Just as the genuine quantum model merges into the unimodal model, so it is possible to find a quantum which fits data from any distribution, since experimental data are usually rational. We have only to take the H.C.F. of the observations, after adding an arbitrary constant, to obtain a quantum which fits the data precisely. Other constants are possible which fit the data less exactly, and even if the observations were not rational, a quantum can be found to fit the data as closely as we please. Equally, when a quantum fits the data we have only Occam's principle to exclude the possibility that the true quantum is a half

or a third of the one we have accepted. The difficulties of estimation in a similar situation have been discussed by Hammersley (1950).

Obviously when so many alternatives are so easily produced we must be careful to state explicitly what assumptions are being made and where independent evidence is being used.

(1·5) In this paper the problems considered will be limited to the following.

(i) Estimation of the positions of the modes when the observations have been allotted to the correct subdistributions, and these subdistributions are normal

(ii) Estimation of the variance of each component when the subdistributions are normal and *either* the observations have been allotted to the correct subdistributions (positions of the modes unknown) *or* the positions of the modes are known (observations not allotted to subdistributions).

(iii) Testing whether to accept a hypothesis which specifies the positions of the modes and which is independent of the data used in the test. The normality of the subdistributions is not here assumed, the p_t are not specified, nor are the observations allotted to subdistributions. The hypothesis against which the quantum hypothesis is compared is one of a class described in § (4·2).

2. ESTIMATION: MODES

(2·1) In this section it is assumed that the data can be allotted to the correct subdistributions, i.e. that when we are given any observation we can say it comes from the tth subdistribution, although we do not know precisely the location of this subdistribution. This may be possible because the data fall into clearly defined groups with every component known to be represented or for some other reason. For example, in the measurement of insects during moulting, each group or subdistribution is composed of measurements after a defined number of moults.

Let y_{rs} be the sth observation in the rth subdivision, with mean $\beta + 2r\delta$. Then

$$y_{rs} = \beta + 2r\delta + \epsilon_{rs},$$

where β and δ are unknown constants (2δ is the quantum), r is zero or a positive integer, and ϵ_{rs} the normal error, or deviation of y_{rs} from its mean, with mean zero. Let $r = 0, 1, ..., m$; $s = 1, 2, ..., n_r$; $\sum_{r=0}^{m} n_r = n$; $\sum_{s=1}^{n_r} y_{rs} = Y_r$.

(2·2) The method that has frequently been used to detect the quantum situation and to estimate the positions of the modes is to plot the means of the successive groups (or the experimental modes; these may also be used when observations cannot be allotted to groups) against r, and to draw a straight line through the points so obtained. A similar but efficient estimation of β and 2δ is by the regression of y_{rs} on r. The method is equivalent to least-squares and to maximum-likelihood procedures for estimating the means of the subdistributions subject to the restrictions that they are in arithmetic progression. The results of the regression analysis are summarized in the three sections below.

(2·3) ϵ_{rs} *has* s.d. σ (*all* r). The equations giving estimates b and $2d$ for β and 2δ are

$$b = (\Sigma r^2 n_r \Sigma Y_r - \Sigma r n_r \Sigma r Y_r)/\Delta,$$

$$2d = (n\Sigma r Y_r - \Sigma r n_r \Sigma Y_r)/\Delta,$$

where $\Delta = n\Sigma r^2 n_r - (\Sigma r n_r)^2$, and all summations are over $r = 0, 1, ..., m$.

The properties deducible from regression coefficients apply to b (the intercept of the regression line with the ordinate) and $2d$ (the slope). We have:

(i) the variance σ^2 of each y_{rs} about its mean is estimated by $s_1^2 = S_1^2/(n-2)$, where

$$S_1^2 = \{\Delta\Delta' - (n\sum_r rY_r - \sum_r rn_r \sum_r Y_r)^2\}/(n\Delta),$$

and

$$\Delta' = n\sum_r\sum_s y_{rs}^2 - (\sum_r Y_r)^2.$$

(ii) S_1^2 has σ^2 times the χ^2 distribution with $(n-2)$ d.f.

(iii) the variance of $2d$, estimator of the quantum 2δ, is itself estimated by $s_1^2\Delta$.

(iv) s_1^2/Δ is independent of $2d$, so the t-test may be applied to hypotheses about 2δ.

(v) The variance of b, estimator of β, is itself estimated (not independently of the variance of $2d$) by

$$s_1^2(1 + (\sum_r rn_r)^2/n\Delta)/n.$$

(2·4) $\beta = 0$ *given*; ϵ_{rs} has S.D. σ (all r). The estimator for 2δ is

$$2d = (\sum_r rY_r)/(\sum_r r^2 n_r).$$

(i) the variance σ^2 of each y_{rs} about its mean is now estimated by $s_2^2 = S_2^2/(n-1)$, where

$$S_2^2 = \sum_r\sum_s y_{rs}^2 - (\sum_r rY_r)^2/\sum_r r^2 n_r.$$

(ii) S_2^2 has σ^2 times the χ^2 distribution with $(n-1)$ d.f.

(iii) the variance of $2d$ is itself estimated by $s_2^2/\sum_r r^2 n_r$.

(iv) $s_2^2/\sum_r r^2 n_r$ is independent of $2d$, so the t-test is applicable to hypotheses about 2δ.

(2·5) $\beta = 0$ *given*; ϵ_{rs} has S.D. $r\sigma$. This is the case in which the S.D. of each observation varies directly with its expected value, i.e. with $2r\delta$ and so with r. In the least-squares equation for estimating 2δ each observation must be weighted inversely by its S.D.; we then obtain the maximum-likelihood estimate

$$2d = \sum_r (Y_r/r)/n.$$

This is the arithmetic mean of the homoscedastic ratios Y_r/r, i.e. in the regression model each observed point gives an equally precise estimate of the slope.

The variance σ^2 is estimated by essentially the same method as that implied in (2·3) and (2·4). Consider the identity

$$(y_{rs} - 2r\delta)/r = (y_{rs} - 2rd)/r + 2(d - \delta).$$

Of these quantities the first is distributed normally with mean zero and variance σ^2, the second does not need δ in its calculation, while the third has mean zero and variance σ^2/n. It follows that

$$\sum_r\sum_s (y_{rs} - 2r\delta)^2/r^2 = \sum_r\sum_s (y_{rs} - 2rd)^2/r^2 + 4n(d - \delta)^2,$$

and by Cochran's theorem that

$$S_3^2 = \sum_r\sum_s (y_{rs} - 2rd)^2/r^2$$

has σ^2 times the χ^2 distribution with $(n-1)$ d.f.

(i) The variance σ^2 is estimated by $s_3^2 = S_3^2 (n-1)$, where S_3^2 is calculated from

$$S_3^2 = \sum_r\sum_s (y_{rs}/r)^2 - \{\sum_r (Y_r/r)\}^2/n,$$

This is otherwise clear, since $y_{rs} - r$ has variance σ^2.

(ii) The variance of $2d$, estimator of 2δ, is itself estimated by s_3^2/n.

(iii) s_3^2 is independent of $2d$, so the t-test may be applied to hypotheses about 2δ.

(iv) If we wish to set limits to a single new observation in the rth group, we must take into account the uncertainty of the supposed mean $2r\delta$, which contributes a variance $r^2\sigma^2/n$, and the variance $r^2\sigma^2$ of a single observation in this group. The limits are

$$r\{2d + t_{n-1}s_3\sqrt{[(n+1)/n]}\},$$

where t_{n-1} denotes the value of Student's t with $(n-1)$ d.f. at the appropriate probability level.

(2·6) *Example of estimation.* Svedberg (1939) gives the molecular weights of fifty-six proteins determined by sedimentation velocity or by sedimentation equilibrium. For twenty proteins both methods have been used so that seventy-six measurements in all are given. Svedberg deduces a 'law of simple multiples....If we choose 17,600 as the unit the majority of the proteins may be divided into eleven classes with molecular weights which are multiples of this unit by factors containing powers of 2 and 3. The rule is only approximate, indicating that the underlying principle is obscured by some secondary factor.' He notes seventy observations in eleven classes which he considers obey this law, and he gives the factor relevant to each class by which the unit is multiplied.

We take these seventy observations as a random sample from a population of the type considered in § (2·5), and for each class take the factor given by Svedberg as r. It is clear from a study of the data that it is appropriate to suppose the S.D. of each class is linearly related to its mean. We can then apply the method of § (2·5) to estimate the quantum for which Svedberg gives the value 17,600. We obtain $2d = 17,920$, $s_3^2 = 3\cdot83$ (69 d.f.), i.e. 95 % limits to 2δ are $17,460 - 18,390$.

Six observations are so far from Svedberg's supposed modes that he did not classify them. When the limits within which a new observation may be expected to lie are calculated it is found that two of the outliers are within 95 % limits, two within 98 % limits, and two within 99 % limits. Therefore these extreme observations are not inconsistent with the remainder of the data.

It may be remarked here that Johnston, Longuet-Higgins & Ogston (1945) consider that Svedberg's data do not support his hypothesis. Their tests will be discussed below in § (4·6).

3. Estimation: variance

(3·1) In §§ (2·3), (2·4) and (2·5) an estimator for σ^2 and the form of its distribution were obtained when the positions of the modes were estimated.

(3·2) Now suppose the positions of the modes are known, independently of the data, but that the observations have not been allotted to subdistributions and may overlap, i.e. lie nearer to another mean than their own. The weights p_t are not known. The distribution is that of § (2·1).

Let z_i $(i = 1, 2, \ldots, n)$ be the distance of y_{rs} from the nearest mode, i.e.

$$z_i = y_{rs} - (\beta + 2r'\delta),$$

where r' is chosen to minimize $|z_i|$. Clearly $|z_i| \leqslant \delta$, and z_i is ϵ_{rs} plus or minus an integer (or zero) multiple of 2δ. For those ϵ_{rs} satisfying $|\epsilon_{rs}| \leqslant \delta$, the distribution of z_i is the same as

4

that of ϵ_{rs}, i.e. normal with mean zero and variance σ^2 truncated at $\pm\delta$. For those ϵ_{rs} satisfying $\delta < \epsilon_{rs} \leqslant 3\delta$, the distribution of z_i is the same as that of $(\epsilon_{rs} - 2\delta)$ truncated at $\pm\delta$, and so on. The effect of transforming from y to z is to cut the overall frequency distribution of y at $\dots \beta - \delta, \beta + \delta, \beta + 3\delta, \dots$ and to lump together the truncated portions. Since each subdistribution is the same, save for its mean and p_t, the result is shown in Fig. 2, where the subdistribution (dotted line) is transformed into the distribution of z (full line) by successively 'turning in' the tails of the distribution. It will be noticed that the p_t do not appear in the distribution of z, nor is the allocation of y_{rs} to its correct subdistribution relevant to z_i.
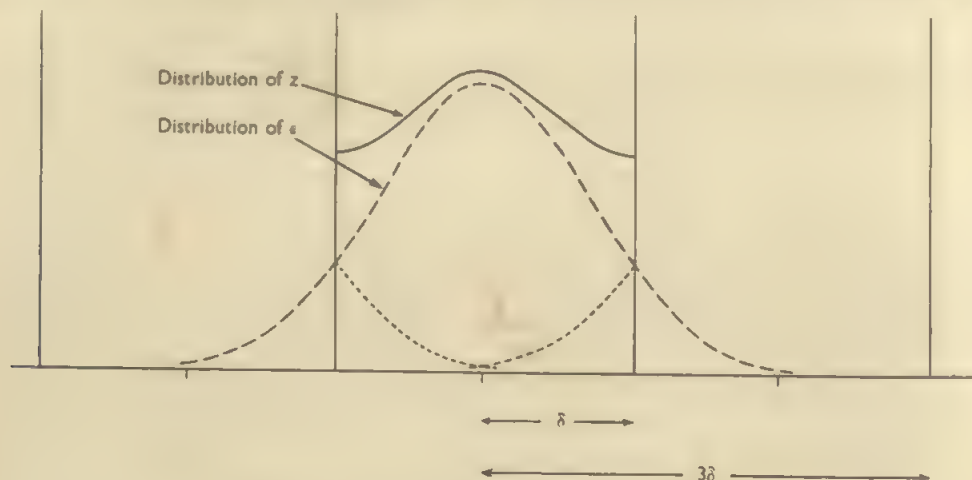
Fig. 2. Lumping the normal distrbution.

The *lumped variance* of the observations, s^2, is defined by

$$s^2 = \sum_{i=1}^{n} z_i^2 / n.$$

We now consider the distribution of s^2/δ^2 on the quantum hypothesis when the sub-distributions are normal. By the central limit theorem s^2/δ^2 is asymptotically normal, and we therefore derive its first two moments. The mean of s^2/δ^2 is the same as the mean of z^2/δ^2, and the variance of s^2/δ^2 is $1/n$ that of z^2/δ^2.

Now each z^2 is equal to ϵ^2 for $|\epsilon| \leqslant \delta$, and for all other ϵ these values are repeated with period 2δ. We may represent z^2 by a Fourier cosine series:

$$z^2 = \tfrac{1}{3}\delta^2 + \frac{4\delta^2}{\pi^2} \sum_{r=1}^{\infty} \frac{(-)^r}{r^2} \cos\left(\frac{r\pi\epsilon}{\delta}\right).$$

When calculating $E[z^2/\delta^2]$ we may integrate the series term-by-term, since it is a uniformly convergent series of continuous terms for all z. We obtain

$$E[z^2/\delta^2] = \frac{1}{3} + \frac{4}{\pi^2} \sum_{r=1}^{\infty} \frac{(-)^r}{r^2} \exp\left(-\frac{r^2\pi^2\sigma^2}{2\delta^2}\right).$$

Expanding z^4 as a Fourier series we obtain similarly

$$E[z^4/\delta^4] = \frac{1}{5} + \frac{8}{\pi^4} \sum_{r=1}^{\infty} \frac{(-)^r}{r^4} (r^2\pi^2 - 6) \exp\left(-\frac{r^2\pi^2\sigma^2}{2\delta^2}\right).$$

The two infinite series thus introduced are functions of $\frac{1}{2}\pi^2\sigma^2 \delta^2$ and have been tabulated by Newman (1934) in this form. They are integrals of one of the theta functions. We give in Table 1 some values of $E[z^2 \delta^2]$ and $\mathrm{Var}[z^2 \delta^2]$ deduced from Newman's tables.

Now we write $E[z^2 \delta^2] = g(\sigma^2 \delta^2)$, the function g being tabulated below. It follows that $s^2 \delta^2$ is an unbiased and consistent estimator of $g(\sigma^2 \delta^2)$.

$$g(\sigma^2/\delta^2) \approx s^2/\delta^2.$$

Since we require an estimator of σ^2 explicitly, we rewrite this equation,

$$\sigma^2 \approx \delta^2 g^{-1}(s^2/\delta^2) \equiv \delta^2 h(s^2/\delta^2).$$

This estimator is consistent and is recommended for general use rather than a maximum-likelihood estimator (which can be calculated if required) because of its simplicity. A table of $h(s^2 \delta^2)$ is needed when this estimator is used; it is given as Table 2, and has been obtained by interpolation in Table 1.

Table 1. *Mean and variance of z^2/δ^2 on the quantum hypothesis.*
The variance of s^2/δ^2 is $1/n$ that of z^2/δ^2

$\frac{1}{2}\pi^2\sigma^2/\delta^2$	$E[z^2, \delta^2]$	$\mathrm{Var}\,[z^2, \delta^2]$	$\frac{1}{2}\pi^2\sigma^2/\delta^2$	$E[z^2\,\delta^2]$	$\mathrm{Var}\,[z^2\,\delta^2]$
0·0	0·0000	0·0000	1·0	0·1861	0·0516
·1	·0203	·0008	1·5	·2431	·0704
·2	·0405	·0033	2·0	·2785	·0795
·3	·0608	·0074	2·5	·3001	·0839
·4	·0809	·0129	3·0	·3132	·0861
0·5	0·1007	0·0194	3·5	0·3211	0·0873
·6	·1199	·0264	4·0	·3259	·0880
·7	·1382	·0334	4·5	·3288	·0883
·8	·1553	·0400	5·0	·3306	·0886
·9	·1713	·0461	∞	·3333	·0889

4. The lumped variance test

(4·1) Suppose we are in the situation of § (3·2), so that the positions of the suspected modes have been given independently of the data. The observations are to be used to test whether there is real grouping about these modes. We neither know the weights p_t to be attached to each mode nor can we allot any observation to its correct subdistribution. The alternative to the quantum hypothesis has not been specified, but we wish to interpret the intuitive feeling that the alternative hypothesis implies no preference for the suspected modes and is perhaps rectangular or unimodal. We could rather not assume too much about the form of the subdistributions (e.g. normality), and a test which is in some sense distribution-free will therefore have advantages.

(4·2) In these circumstances a test using the lumped variance is appropriate. It has already been noted that neither the p_t nor the allocation of observations to the correct subdistributions are relevant to the distribution of s^2. Further, when the observations are clustered about the suspected modes s^2 is small in relation to δ^2. The tendency of the observations to be grouped at the modes is therefore measured by s^2/δ^2, and the form of the

subdistributions will be used only in this respect. It is, of course, true that if the true modes are not those suspected but are near them, and σ/δ is small, then s^2/δ^2 will again be small.

If y has a rectangular distribution, z has also a rectangular distribution between $-\delta$ and $+\delta$. If y has a smooth unimodal distribution whose spread is not small in comparison with δ, a little thought will show that z has approximately the rectangular distribution. Therefore, we take the hypothesis that z is rectangularly distributed between $-\delta$ and $+\delta$ to be our null hypothesis. On this rectangular hypothesis s^2/δ^2 is the mean of n independent variates each of which can be shown to have mean $\frac{1}{3}$ and variance $\frac{4}{45}$. These are also the limits of the mean and variance of z^2/δ^2 on the quantum hypothesis as $\sigma/\delta \to \infty$. By the central limit theorem, s^2/δ^2 is approximately normally distributed with mean $\frac{1}{3}$ and variance $\frac{4}{45}n$ when n is large; for $n \geqslant 20$ the approximation may be expected to be good.

Table 2. $h(s^2/\delta^2)$. *An estimate of* σ^2 *is given by* $\delta^2 h(s^2/\delta^2)$

s^2/δ^2	$h(s^2/\delta^2)$	s^2/δ^2	$h(s^2/\delta^2)$
0	0·000	0·16	0·167
0·01	·010	·17	·180
·02	·020	·18	·194
·03	·030	·19	·208
·04	·040	·20	·223
0·05	0·050	0·21	0·238
·06	·060	·22	·255
·07	·070	·23	·274
·08	·080	·24	·295
·09	·090	·25	·318
0·10	0·100	0·26	0·344
·11	·111	·27	·374
·12	·122	·28	·409
·13	·132	·29	·452
·14	·143	·30	·506
·15	·155	·31	·577
		·32	·690

On the quantum hypothesis s^2/δ^2 has expectation less than $\frac{1}{3}$. As σ^2 increases the distribution, although grouped, becomes indistinguishable from one of the class of alternatives.

(4·3) The lumped variance test is a one-sided test of the rectangular hypothesis, and gives the probability that the value of s^2/δ^2 found, or a lower value, would occur by chance if z were rectangularly distributed. It will be noted that it is the alternative, and not the quantum hypothesis which is tested.

The significance points, for probability levels 0·05, 0·01 and 0·001 and $n = 20(5) 100(50)$ 1000 are given in Table 3. If the value of s^2/δ^2 found is less than the appropriate value in this table, the rectangular hypothesis is rejected and the hypothesis of grouping, or quantum hypothesis, is accepted. It will be remembered that the rectangular hypothesis may be accepted if the variance σ^2 of the subdistributions is large in relation to δ^2.

(4·4) *Example: distribution of length in* Labidocera euchaeta. Seymour Sewell (1912) measured the total lengths of large numbers of Copepoda with a view to testing whether

they followed Brook's law (Fowler, 1909) as stated for Somatopoda and Ostracoda. From his data we extract Table 4, of the distribution of length of females of the species *Labidocera euchaeta*. Sample A (Seymour Sewell's fig. 1) consists of 497 specimens from the Rangoon

Table 3. *Lumped variance test. If $s^2 \, \delta^2$, calculated from n observations, is less than the value tabulated at probability level P, the quantum hypothesis is accepted at that probability level*

n	P = 0·05	P = 0·01	P = 0·001
20	0·2237	0·1782	0·1273
25	·2353	·1946	·1490
30	·2438	·2067	·1651
35	·2504	·2161	·1776
40	·2558	·2237	·1878
45	0·2602	0·2299	0·1960
50	·2640	·2352	·2030
55	·2672	·2398	·2091
60	·2700	·2438	·2144
65	·2725	·2473	·2191
70	0·2747	0·2504	0·2232
75	·2767	·2532	·2269
80	·2785	·2558	·2303
85	·2801	·2581	·2334
90	·2816	·2602	·2362
95	·2830	·2622	·2388
100	0·2843	0·2640	0·2412
150	·2933	·2767	·2581
200	·2987	·2843	·2682
250	·3023	·2895	·2751
300	·3050	·2933	·2801
350	0·3071	0·2963	0·2841
400	·3088	·2987	·2873
450	·3102	·3006	·2899
500	·3114	·3023	·2921
550	·3124	·3038	·2940
600	0·3133	0·3050	0·2957
650	·3141	·3061	·2972
700	·3148	·3071	·2985
750	·3154	·3080	·2997
800	·3160	·3088	·3008
850	0·3165	0·3095	0·3017
900	·3170	·3102	·3026
950	·3174	·3108	·3034
1000	·3178	·3114	·3042

River Estuary, and sample B (his fig. 2) of 157 collected after sample A from Chittagong. We shall use these data to answer the question 'Do the specimens from Chittagong show any tendency to occur relatively more frequently at the modes from the first sample?'

It is expected from previous work that when length is taken on a logarithmic scale the data will be grouped about approximately equally spaced modes (any change in the growth factor is here being neglected). Because the groups overlap and no other information is available, the positions of the population modes of sample A cannot be estimated by the methods of § (2). When the experimental modes at 15, 22, 30, 42, 54 and 65 units of length (the modes at 20, 36 and 49 units are seen from a figure to be spurious) are plotted on a

Table 4. *Distribution of lengths of females.* Labidocera euchaeta
(*unit of length* 0·04 mm.). *Semour Sewell's data*

Length	Sample A	Sample B	Length	Sample A	Sample B
13	2	0	43	15	6
14	3	0	44	5	3
15	3	0	45	6	3
16	3	0	46	9	2
17	1	0	47	16	2
18	0	0	48	18	2
19	1	0	49	23	2
20	3	0	50	21	8
21	2	5	51	25	11
22	·4	0	52	25	12
23	1	0	53	28	10
24	2	0	54	31	9
25	4	0	55	19	9
26	5	1	56	9	7
27	10	2	57	4	1
28	10	5	58	2	2
29	12	4	59	1	0
30	14	2	60	0	1
31	12	0	61	1	0
32	8	0	62	0	1
33	5	0	63	1	0
34	1	1	64	7	2
35	2	1	65	12	2
36	6	3	66	9	0
37	5	4	67	7	1
38	7	7	68	4	0
39	12	6	69	3	0
40	18	5	70	1	1
41	18	6	71	1	1
42	20	7			
			Totals	497	157

logarithmic scale against $r = 0, 1, \ldots, 5$ they lie approximately on a straight line. Fitting a line by eye we conclude that sample A suggests modes at 1·218 (0·126) 1·848 on a log-arithmic scale.

We next require to test whether sample B supports this grouping, and do so by the lumped variance test. Taking logarithms, we calculate the square of the distance of each observation of sample B from the nearest mode deduced from sample A. The sum of these squared distances, divided by $n\delta^2 = 157 (0·126 2)^2$, is 0·177. Comparison of this value with Table 3 shows that departure from the value expected on the rectangular hypothesis is highly

significant. We conclude that sample B supports strongly the grouping we have deduced from sample A, i.e. that grouping is real and occurs near the modes given above.

If an estimate of the scatter about these modes is required, and the assumptions of § (3·2) are justified, the method of that section gives an estimate of the variance about each mode on a logarithmic scale. Since s^2/δ^2 is 0·177, and δ^2 is $(0·063)^2$, the variance is estimated as

$$\delta^2 h(s^2/\delta^2) = (0·063)^2 (0·189) = 0·000750.$$

(4·5) *Example: choosing a random point.* The following simple experiment is described because it exemplifies the use of lumped variance. It was designed to test whether the points on a line chosen 'at random' by a number of people were in fact grouped about certain regular points on the line.

Twenty subjects were chosen from workers at B.C.U.R.A. and each was presented with a sheet of paper on which were drawn eight parallel lines of length 12 cm. Each subject was
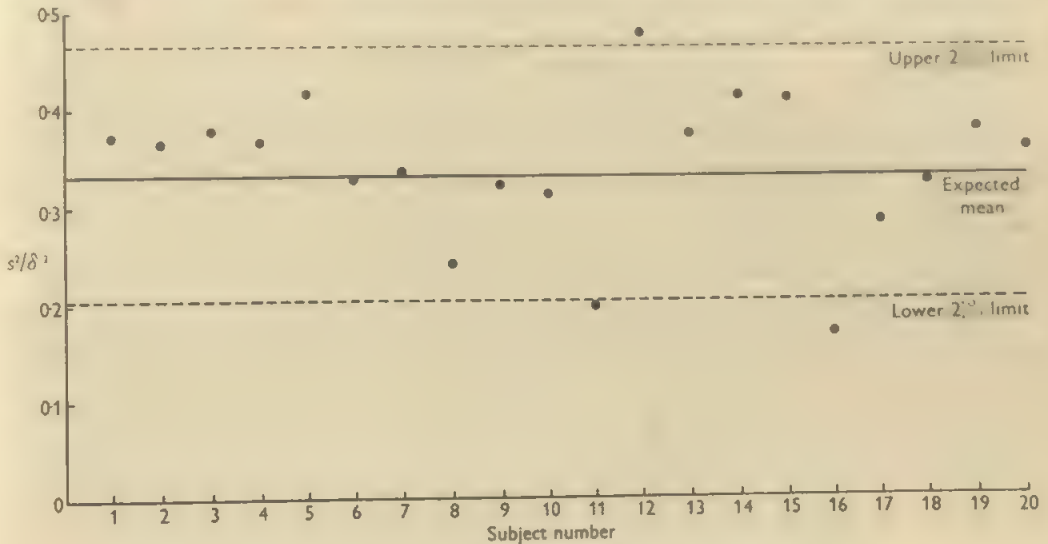
Fig. 3. Choosing a random point. Control chart of s^2/δ^2.

then asked to mark points on the lines at random, i.e. without systematic placing. He was to make one mark on each of the first two lines, two, three and four on each of the successive pairs of lines, making twenty marks in all. Alternate subjects marked their lines in the reverse order.

The points about which systematic grouping was expected were those which divide the line into $(n + 1)$ equal segments when n marks were made. Thus when one mark was made it was expected the end-points and centre would be preferred, when two marks were made the end-points and thirds, and so on. In this situation the lumped variance test is applicable. When n points are marked and l is the length of the line, δ is $l/2(n + 1)$, and z is the distance of a mark from the nearest 'preferred' point as defined above.

The values for s^2/δ^2 found in the experiment are given in Table 5, and are shown also for subjects in the form of a control chart. In Table 5, $1a$ denotes the first line on which one mark was made, $1b$ the second line on which one was made, and so on.

The conclusion drawn is that no evidence was found that the subjects did not choose points on the lines at random; the data are consistent with the hypothesis that all points

on the line are equally likely to be chosen. Clearly more extensive experiments and other tests may reverse this conclusion. One subject has $s^2 \delta^2$ at about the 99 % point, i.e. his marks tend to avoid the regular points, and two subjects have $s^2 . \delta^2$ below the 3 % point, i.e. their marks do cluster at these points, but these are the extremes in a sample of twenty. None of the $s^2 . \delta^2$ for lines, nor the general average, is significantly different from $\frac{1}{3}$ as expected on the rectangular hypothesis. Further, the variance of $s^2 \delta^2$ for subjects is 0·0056; this does not differ significantly from the variance expected on the rectangular hypothesis, 0·0044.

(4·6) Finally, we consider the statistical tests applied by Johnston *et al.* (1945) to Svedberg's (1939) data. We have already noted, in § (1·2), the difficulty in testing a hypothesis which is suggested by the data used in the test. Suppose, however, that a lumped variance

Table 5. *Choosing a random point*

Subject number	s^2/δ^2	Subject number	s^2/δ^2	Line	s^2/δ^2
1	0·371	11	0·199	1a	0·369
2	·364	12	·479	1b	·294
3	·379	13	·373		
4	·366	14	·414	2a	0·345
5	·417	15	·413	2b	·316
6	0·328	16	0·170	3a	0·360
7	·336	17	·286	3b	·309
8	·241	18	·325		
9	·321	19	·380	4a	0·348
10	·311	20	·359	4b	·364
		Average	0·342	Weighted average	0·342

test is applied to such data, using the modes suggested by or calculated from the data, and that the value of s^2/δ^2 found is not significantly less than the value $\frac{1}{3}$ expected on the rectangular hypothesis. Then we would correctly reject the quantum hypothesis and accept the class of distribution defined in § (4·2). The reason for doing so is that when the modes are suggested by the data the value of s^2/δ^2 will therefore be small, corresponding to the grouping of the data about these modes. In such circumstances it would be wrong to accept the quantum hypothesis solely because s^2/δ^2 is less than a conventional significance point, whereas it is proper to reject the quantum hypothesis if s^2/δ^2 does not reach significance.

The 'correlation function' test applied by Johnston *et al.* is similar to the lumped variance test proposed in § (4·3); their statistic stands in much the same relation to the lumped variance as the mean deviation does to the variance. The lumped variance has the advantage that it is useful in the estimation of variance by the method of § (3·2). Their statistic is, in the notation defined above,

$$F(n) = 1 - 2 \sum_{i=1}^{n} \{|z_i|/(n\delta)\}.$$

It is distributed approximately normally about zero with variance $1/(3n)$ on the rectangular hypothesis, and a positive value indicates grouping about the supposed modes.

They extend this definition of $F(n)$ to those cases in which the modes are not equally spaced, taking as the δ corresponding to each z_i half the distance between the modes on either side of z_i. A similar extension is, of course, possible for the statistic $\sum_{i=1}^{n} z_i^2/(n\delta^2)$ of §(3·2). The rectangular hypothesis they specify in this case to be a rectangular distribution of y between each mode; this is unnecessarily restricted, for we can consider the class of all distributions of y which correspond to a rectangular distribution of z.

Johnston *et al.* apply their test to the fifty-six values given by Svedberg, averaging the results of two determinations when these are available. They test for grouping about the values $2^n 3^m \times 17,600$ for all integral n and m, and obtain for $F(n)$ the value $0·137$ with S.D. $0·077$; this corresponds to a two-tailed probability of $0·076$. Since this does not reach significance they conclude Svedberg's hypothesis receives no support from the evidence. A subdivision of the data is intended to show that only observations near the first two Svedberg numbers support his hypothesis. Two further tests, applied to part of the data only, and whose power may be suspected to be low, suggest a similar conclusion.

The probability given by Johnston *et al.* should properly be one-tailed, since the alternative hypothesis suggests a positive $F(n)$. Their value therefore exceeds the 5% significance level. Moreover, Svedberg proposed eleven groups, the factor multiplying 17,600 being of the form $2^n 3^m$ for certain values of n and m only. When the 'correlation function' test is applied to the values given by Svedberg, $F(n)$ exceeds the $0·1$% significance level. The lumped variance test gives the same result, indicating, contrary to the findings of Johnston *et al.*, that the data do support Svedberg's hypothesis. For the reasons already given (§1·2) no positive conclusions may be drawn from this.

REFERENCES

BROOKS, W. K. (1886). *Report on the Scientific Results of the Exploring Voyage of H.M.S. Challenger, 1873–76*, **16**, (Zoological Section), *Stomatopoda*, 1.

FOWLER, G. H. (1909). *Trans. Linn. Soc. (Zool.)*, 2nd ser. pt. 9.

GRANT, P. J. (1952). *Proc. Phys. Soc., Lond.*, A, **65**, 150.

HAMMERSLEY, J. M. (1950). *J. R. Statist. Soc.* B, **12**, 192.

HAMMERSLEY, J. M. & MORTON, K. W. (1954). *J. R. Statist. Soc.* B, **16**, 23.

JOHNSTON, J. P., LONGUET-HIGGINS, H. C. & OGSTON, A. G. (1945). *Trans. Faraday Soc.* **41**, 588.

NEWMAN, A. B. (1934). *Trans. Amer. Inst. Chem. Engrs*, **30**, 598.

PEARSON, E. S. & CHANDRASEKHAR, C. (1936). *Biometrika*, **28**, 308.

PEARSON, K. (1894). *Phil. Trans.* A, **185**, 71.

PRZIBRAM, H. & MEGUŠAR, F. (1912). *Roux Arch. EntwMech. Organ.* **34**, 680.

SEYLER, C. A. (1943). *J. Inst. Fuel*, **16**, 134.

SEYMOUR SEWELL, R. B. (1912). *Rec. Indian Mus.* **7**, 313.

SVEDBERG, T. (1939). *Proc. Roy. Soc.* A, **170**, 40 and B, **127**, 1.

THOM, A. (1955). *J. R. Statist. Soc.* B. (in the Press).

WIGGLESWORTH, V. D. (1942). *Principles of Insect Physiology*, 2nd ed. London: Methuen.

THE TRUNCATED NEGATIVE BINOMIAL DISTRIBUTION

By M. R. SAMPFORD

Agricultural Research Council Unit of Statistics, University of Aberdeen

1. INTRODUCTION

The negative binomial distribution has been discussed by, *inter alia*, Greenwood & Yule (1920), Fisher (1941), Haldane (1941), Anscombe (1950) and Bliss & Fisher (1953), and is extensively used for the description of data too heterogeneous to be fitted by a Poisson distribution. Observed samples, however, may be truncated, in the sense that the number of individuals falling into the zero class cannot be determined. For example, if chromosome breaks in irradiated tissue can occur only in those cells which are at a particular stage of the mitotic cycle at the time of irradiation, a cell can be demonstrated to have been at that stage only if breaks actually occur. Thus in the distribution of breaks per cell, cells not susceptible to breakage are indistinguishable from susceptible cells in which no breaks occur.

Methods for estimation of the parameters of the truncated distribution are considered in this paper. The corresponding problem of estimation of the truncated Poisson distribution has been discussed by David & Johnson (1952), who also discuss the present problem.

2. THE MOMENTS OF THE TRUNCATED DISTRIBUTION

The negative binomial distribution has the form (in Fisher's notation)

$$P(r) = \frac{(k+r-1)!}{(k-1)!\,r!}\frac{p^r}{(1+p)^{k+r}} \quad (r = 0, 1, \ldots; \; p, k > 0), \tag{1}$$

so that
$$P(0) = 1/(1+p)^k.$$

To obtain the corresponding probabilities for the truncated distribution the form (1) must be divided by $(1 - P(0))$; writing

$$\varpi = 1/(1+p), \quad \eta = 1 - \varpi,$$

it follows that

$$P_t(r) = \frac{\varpi^k}{1-\varpi^k}\frac{(k+r-1)!}{(k-1)!\,r!}\,\eta^r \quad (r = 1, 2, \ldots). \tag{2}$$

The factorial moments of this distribution are

$$\mu'_{[j]} = \frac{(k+j-1)!\,\eta^j}{(k-1)!\,\varpi^j(1-\varpi^k)}, \tag{3}$$

whence
$$\mu'_1 = \frac{k\eta}{\varpi(1-\varpi^k)},$$

$$\mu'_2 = \frac{k\eta + k^2\eta^2}{\varpi^2(1-\varpi^k)},$$

$$\mu'_3 = \frac{k\eta + k(3k+1)\,\eta^2 + k^3\eta^3}{\varpi^3(1-\varpi^k)},$$

$$\mu'_4 = \frac{k\eta(6 - 6\varpi + \varpi^2) + k^2\eta^2(11 - 4\varpi) + 6k^3\eta^3 + k^4\eta^4}{\varpi^4(1-\varpi^k)}.$$

3. ESTIMATION BY MOMENTS

David & Johnson (1952) suggest that the use of estimates of less than maximum efficiency is justifiable only if they are directly obtainable as explicit solutions of easily constructed equations. In discussing the truncated negative binomial, therefore, they do not consider estimates based on the first two sample moments, which do not provide explicit solutions, but confine their attention to a method using certain ratios of the first three moments. i.e. sample estimates of μ'_2/μ'_1 and μ'_3/μ'_1. The estimates (only that of p is discussed in detail) are obtained easily enough, but, in consequence of the introduction of the third sample moment, are extremely inefficient. (For example, the efficiency for values of the parameters equivalent to $k = 1$, $\varpi = 0.5$ is as low as 1.7%.) David & Johnson therefore abandon completely the use of moments and recommend the maximum-likelihood method for use in all cases.

Whether, in fact, any particular inefficient procedure is acceptable can only depend on the loss of information resulting from its use, the time saved by it, and the relative costs of the time or labour spent on observation and on analysis. Thus if an experiment involves observations on hundreds of experimental animals, made over a period of several years, ten or even a hundred hours of calculation may be dearly saved at the cost of 10% of the information so laboriously accumulated. If, on the other hand, the observations are made easily and at no great cost, the use of a convenient but statistically 'inefficient' method of analysis, coupled with an appropriate increase in sample size, may be far more 'efficient' than a tedious maximum-likelihood calculation, in the sense of giving the same amount of information at a lower cost. In this section I give a trial-and-error method for solving the moment equations using the first two moments. This method, though not explicit, does not take many minutes to carry through. It certainly entails far less labour than the maximum-likelihood calculations, and the estimates obtained have a percentage efficiency of 80 or upwards for all but the most unfavourable combinations of parameters: for $k = 1$, $\varpi = 0.5$ the efficiency is 77.5% (§ 6, Table 2), as compared with the 1.7% of the 'three-moment' estimate. This method can thus be recommended as a reasonable alternative to the maximum-likelihood calculations, in circumstances where a method of less than 100% efficiency seems likely to prove acceptable.

Equating the population mean and variance to the corresponding sample values gives equations for k and ϖ:

$$\left.\begin{aligned}
\frac{k\eta}{\varpi(1-\varpi^k)} &= m, \\
\frac{k\eta(1+k\eta)}{\varpi^2(1-\varpi^k)} &= m^2+s^2.
\end{aligned}\right\} \tag{4}$$

These equations can be solved by trial and error, for which purpose they are most conveniently expressed in the form

$$\left.\begin{aligned}
k &= \left(m+\frac{s^2}{m}-1\right)\frac{\varpi}{1-\varpi}-1, \\
\xi &= m\varpi^{k+1}+\frac{s^2}{m}\varpi-1 = 0.
\end{aligned}\right\} \tag{5}$$

The value of ξ can be evaluated for selected values of ϖ, and a solution, if one exists, reached by successive linear interpolations.

To investigate whether a solution exists, and to simplify still further the computations required, we consider the function

$$\phi(x) = \frac{-x \log_e x}{1-x}. \tag{6}$$

This function is tabulated, for values of x between 0 and 1, in Table 4, and can be shown to possess the properties

$$\phi(x) \to 0, 1 \quad \text{as} \quad x \to 0, 1,$$

$$\phi'(x) > 0, \quad \phi''(x) < 0, \quad 0 < x < 1.$$

The second of equations (5) then takes the form

$$\zeta(\varpi) = m \exp\left\{-\left(m + \frac{s^2}{m} - 1\right)\phi(\varpi)\right\} + \frac{s^2}{m}\varpi = 1. \tag{7}$$

If all the observed values equal 1, $m = 1$ and $s^2 = 0$, and $\zeta \equiv 1$. Thus in this situation the moment estimation procedure, naturally enough, fails. In all further discussion, therefore, m will be assumed > 1.

It is easily verified that

$$\varpi_0 = 1 \bigg/ \left(m + \frac{s^2}{m}\right) \quad (<1)$$

satisfies equation (7). However, the corresponding value of k, obtained from the first of equations (5), is 0. ϖ_0, therefore, and any lower values of ϖ (corresponding to negative values of k) are inadmissible: we require a solution of (7) in the range

$$\varpi_0 < \varpi < 1.$$

We have $\zeta(0) = m > 1, \quad \zeta(\varpi_0) = 1, \quad \zeta'' > 0 \quad (0 < \varpi < 1).$

(The result on ζ'' follows from that on ϕ'' quoted above.) There exists, therefore, at most one solution of (7) other than ϖ_0 in the range $0 < \varpi < 1$. The condition that there shall be another solution less than 1 is

$$\zeta(1) > 1,$$

the condition that it shall be greater than ϖ_0 is

$$\zeta'(\varpi_0) < 0.$$

These inequalities reduce to limitations on the relative magnitudes of m and s^2, most conveniently expressed in the form

$$1 - \exp\left\{-\left(m + \frac{s^2}{m} - 1\right)\right\} < \left(m + \frac{s^2}{m} - 1\right)\bigg/ m < \log\left(m + \frac{s^2}{m}\right). \tag{8}$$

Neither of these inequalities is necessarily satisfied; it is quite possible to construct samples in which they are not. In particular, the lower inequality is not satisfied if $s^2 = 0$, whatever the value of m (reasonably enough; we should hardly expect to get sensible estimates by equating an essentially positive function to zero). However, provided very small samples are avoided, and provided m is not too near 1, (8) will probably be satisfied.

Equation (7) can be solved either iteratively, writing

$$\varpi_{i+1} = \frac{m}{s^2}\left[1 - m\exp\left\{-\left(m + \frac{s^2}{m} - 1\right)\phi(\varpi_i)\right\}\right].$$

(9)

or by trial and error. Inasmuch as the solution must be on that part of the ζ curve for which $\zeta' > 0$, between the (unknown) minimum of ζ and the point $\varpi = 1$, it is most convenient to take as the first trial value of ϖ either 1, or, if it is less than 1, $\varpi = m/s^2$ (clearly $\zeta > 1$ for this value), and to work down to the solution. This approach will ensure the convergence of the iterative procedure (9), and should minimize the amount of labour required for the trial-and-error procedure. In the latter case time will probably be saved by calculating, for the first trial value of ϖ, the slope

$$\zeta'(\varpi) = \frac{s^2}{m} - m\left(m + \frac{s^2}{m} - 1\right)\left[\frac{\phi(\varpi) - \varpi}{\varpi(1 - \varpi)}\right]\exp\left\{-\left(m + \frac{s^2}{m} - 1\right)\phi(\varpi)\right\}.$$

(10)

Except for the function in square brackets ($= \phi'(\varpi)$) all the terms used in calculating this expression will already have been evaluated in the calculation of ζ. This slope can be used in choosing the second trial value, after which one or two linear interpolations or extrapolations should be sufficient. If $\varpi = 1$ is taken as the initial approximate, $\phi'(\varpi)$ should be replaced by its limiting value of $\frac{1}{2}$.

4. EXAMPLE OF ESTIMATION BY MOMENTS

In an investigation into chromosome breakage, the following sample distribution of breaks per cell was obtained:

$$r = 1(11),\ 2(6),\ 3(4),\ 4(5),\ 6(1),\ 8(2),\ 9,\ 11,\ 13.$$

$$n = 32,\quad \Sigma r = 110,\quad \Sigma r^2 = 686.$$

$$m = 3\cdot4375,\quad s^2 = 9\cdot9315,\quad s^2/m = 2\cdot8892.$$

Equation (7) is thus

$$3\cdot4375\,e^{-5\cdot3267\phi} + 2\cdot8892\varpi = 1.$$

Taking as the first trial value

$$\varpi = 1/2\cdot8892 = 0\cdot3461,$$

we obtain from Table 4

$$\phi = 0\cdot5616,$$

whence

$$(\phi - \varpi)/\varpi(1 - \varpi) = 0\cdot9522.$$

Then

$$5\cdot3267\phi = 2\cdot9915,$$

and from tables of the negative exponential function, or of natural logarithms,

$$e^{-2\cdot9915} = 0\cdot05021,\quad 3\cdot4375\,e^{-2\cdot9915} = 0\cdot1726,$$

whence

$$\zeta = 0\cdot1726 + 1\cdot0000 = 1\cdot1726$$

and

$$\zeta' = 2\cdot8892 - (0\cdot1726 \times 5\cdot3267 \times 0\cdot9522) = 2\cdot0138,$$

suggesting, as the next trial value,

$$\varpi = 0\cdot3461 - 0\cdot1726/2\cdot0138 = 0\cdot2604.$$

The remainder of the trial-and-error solution, shown in Table 1 (*a*), leads to the value

$$\varpi = 0.2346,$$

whence

$$k = 5.3267 \times 0.2346/0.7654 - 1 = 0.633.$$

Alternatively, by the iterative method, starting from the same initial value, the second approximation

$$\varpi_2 = (1 - 0.1726)/2.8892 = 0.2864.$$

The remainder of the calculations, leading after sixteen cycles to the same solution, are shown in Table 1 (*b*).

Table 1. *Solution of the moment estimation equation for the example of* § 4

(*a*) By trial and error

ϖ	ϕ	$3.4375\,e^{-5.3267\phi}$	ζ	ζ'
0.3461	0.5616	0.1726	1.1726	2.0137
0.2604	0.4737	0.2757	1.0280	
0.2438	0.4550	0.3046	1.0090	
0.2359	0.4459	0.3197	1.0013	
0.2346	0.4444	0.3222	1.0000	

(*b*) By iteration

ϖ	ϕ	$3.4375\,e^{-5.3267\phi}$
0.3461	0.5616	0.1726
0.2864	0.5018	0.2374
0.2639	0.4776	0.2700
0.2527	0.4651	0.2886
0.2462	0.4578	0.3000
0.2423	0.4533	0.3073
0.2397	0.4503	0.3123
0.2380	0.4484	0.3154
0.2370	0.4472	0.3175
0.2362	0.4463	0.3190
0.2357	0.4457	0.3200
0.2353	0.4452	0.3209
0.2350	0.4448	0.3216
0.2348	0.4446	0.3219
0.2347	0.4445	0.3221
0.2346	0.4444	0.3222
0.2346		

(Intermediate stages of the calculations are shown in Table 1 for the sake of the example. In practice either calculation can be carried through on a desk computer, recording only the successive values of ϖ and, for the trial-and-error process, ζ.)

Four decimal places have been retained in the estimation of ϖ as a demonstration of the degree of precision to which the moment equations can be solved by this method. In fact, for this example the variances (§ 5) are so large that there is no practical point in retaining more than two decimal places for ϖ and one for k.

5. THE VARIANCES OF THE MOMENT ESTIMATES

The asymptotic variances and covariances of the moment estimates of ϖ and k are given by the formulae

$$J^2 V(\varpi) = \left(\frac{\partial \mu_2'}{\partial k}\right)^2 V(m_1') - 2\frac{\partial \mu_1'}{\partial k}\frac{\partial \mu_2'}{\partial k}\operatorname{cov}(m_1', m_2') + \left(\frac{\partial \mu_1'}{\partial k}\right)^2 V(m_2').$$

$$J^2 \operatorname{cov}(\varpi, k) = -\frac{\partial \mu_2'}{\partial k}\frac{\partial \mu_2'}{\partial \varpi} V(m_1') + \left(\frac{\partial \mu_1'}{\partial k}\frac{\partial \mu_2'}{\partial \varpi} + \frac{\partial \mu_1'}{\partial \varpi}\frac{\partial \mu_2'}{\partial k}\right)\operatorname{cov}(m_1', m_2') - \frac{\partial \mu_1'}{\partial k}\frac{\partial \mu_1'}{\partial \varpi} V(m_2').$$

$$J^2 V(k) = \left(\frac{\partial \mu_2'}{\partial \varpi}\right)^2 V(m_1') - 2\frac{\partial \mu_1'}{\partial \varpi}\frac{\partial \mu_2'}{\partial \varpi}\operatorname{cov}(m_1', m_2') + \left(\frac{\partial \mu_1'}{\partial \varpi}\right)^2 V(m_2').$$

where

$$\frac{\partial \mu_1'}{\partial k} = \frac{\mu_1'}{k}\{1 - \phi(\varpi^k)\},$$

$$\frac{\partial \mu_1'}{\partial \varpi} = -\frac{\mu_1'}{\varpi\eta}\{1 - \varpi^{k+1}\mu_1'\},$$

$$\frac{\partial \mu_2'}{\partial k} = -\frac{\mu_1'}{k\varpi} + \frac{\mu_2'}{k}\{2 - \phi(\varpi^k)\},$$

$$\frac{\partial \mu_2'}{\partial \varpi} = \frac{\mu_1'}{\varpi\eta} - \frac{\mu_2'}{\varpi\eta}\{2 - \varpi^{k+1}\mu_1'\},$$

$$J = \frac{\partial \mu_1'}{\partial k}\frac{\partial \mu_2'}{\partial \varpi} - \frac{\partial \mu_1'}{\partial \varpi}\frac{\partial \mu_2'}{\partial k}.$$

and $\quad nV(m_1') = (\mu_2' - \mu_1'^2), \quad n\operatorname{cov}(m_1', m_2') = (\mu_3' - \mu_1'\mu_2'), \quad nV(m_2') = (\mu_4' - \mu_2'^2).$

Inserting into these formulae the moment estimates obtained in the example of the previous section,
$$\varpi = 0\cdot 2346, \quad k = 0\cdot 633,$$

we have $(\log_e \varpi = -1\cdot 44988, \varpi^k = 0\cdot 3994)$:

$$\mu_1' = 3\cdot 438578, \quad \mu_2' = 21\cdot 758578, \quad \mu_3' = 215\cdot 773952, \quad \mu_4' = 2941\cdot 290671,$$

whence $\quad V(m_1') = 0\cdot 31046123, \quad \operatorname{cov}(m_1', m_2') = 4\cdot 4048558, \quad V(m_2') = 77\cdot 120467.$

Also $\quad \partial\mu_1'/\partial k = 2\cdot 1164, \quad \partial\mu_1'/\partial\varpi = -12\cdot 9798, \quad \partial\mu_2'/\partial k = 24\cdot 6106, \quad \partial\mu_2'/\partial\varpi = -184\cdot 1591$

$$J = -70\cdot 31365336,$$

whence $\quad V(\varpi) = 0\cdot 015091, \quad \operatorname{cov}(\varpi, k) = 0\cdot 08125, \quad V(k) = 0\cdot 4983.$

6. THE EFFICIENCY OF THE MOMENT PROCEDURE

It has been shown by Fisher (1941) that the method of moments may be seriously inefficient for the estimation, from complete data, of the parameters of the negative binomial distribution. Investigation of the corresponding efficiency for estimation from truncated data therefore seems desirable.

The determinant of the variance-covariance matrix of the moment estimates reduces to

$$\frac{2(k+1)\varpi^2(1-\varpi^k)^3}{n^2\eta} \times \frac{\{1 - \varpi^k[1 + k\eta + \frac{1}{2}k(k+1)\eta^2]\}}{\{1 - \varpi^k[1 - k^2\eta - k(k+1)\log\varpi]\}^2}.$$

The determinant of the information matrix for maximum efficiency is

$$\frac{n^2\{1 - \varpi^k[1 + k\eta]\}}{\varpi^2\eta(1-\varpi^k)^3}\left\{\sum_{r=2}^{\infty}\frac{\eta^r}{r}\frac{(r-1)!\,k!}{(k+r-1)!} - \frac{k\varpi^k(\eta + \log\varpi)^2}{\{1 - \varpi^k[1 + k\eta]\}}\right.$$

whence the reciprocal of the efficiency of the moment method, given by the product of these two expressions, is

$$\frac{1}{E} = \frac{\{1 - \varpi^k[1 + k\eta + \tfrac{1}{2}k(k+1)\,\eta^2]\}\,\{1 - \varpi^k[1 + k\eta]\}}{\{1 - \varpi^k[1 - k^2\eta - k(k+1)\log\varpi]\}^2}$$

$$\times \left\{ 2 \sum_{r=2}^{\infty} \frac{\eta^{r-2}\,(r-1)!\,(k+1)!}{r \quad (k+r-1)!} - \frac{2k(k+1)\,\varpi^k}{[1 - \varpi^k(1 + k\eta)]}\left[1 + \frac{\log\varpi}{\eta}\right]^2 \right\}.$$

Table 2 shows the percentage efficiency of the moment method for selected values of k, and of the mean of the complete distribution. From this it is clear that even for quite small means there may be a serious loss of efficiency for low values of k.

For the example of § 4, k was estimated as 0·633, and the mean of the complete distribution $= k\eta/\varpi = 2.065$, from which estimation by moments would appear to be about 70 % efficient in this case.

Table 2. *Percentage efficiency of the moment method of estimation, for selected values of k and of the mean of the complete distribution*

Mean \ k	0·5	1	2	3	4	5
0·5	82·4	90·6	95·7	97·5	98·3	98·8
1·0	74·6	84·7	92·4	95·4	96·9	97·8
2·0	66·1	77·5	87·5	92·0	94·4	95·9
5·0	55·9	67·3	79·2	85·5	89·2	91·6

7. ESTIMATION BY MAXIMUM LIKELIHOOD

In view of the results of the previous section, it seems desirable to consider the maximum-likelihood estimation procedure in some detail.

The log likelihood is

$$\log L = \sum_{1}^{\infty} n_r \log \left\{ \frac{\varpi^k}{1 - \varpi^k} \frac{(k+r-1)!}{(k-1)!\,r!}\,(1-\varpi)^r \right\}$$

$$= nk\log\varpi - n\log(1-\varpi^k) + \sum_{1}^{\infty} rn_r \log(1-\varpi) - \sum_{1}^{\infty} n_r \log r! + \sum_{r=1}^{\infty} n_r \sum_{j=1}^{r} \log(k+j-1),$$

giving the maximum-likelihood equations

$$\frac{nk}{\varpi(1-\varpi^k)} - \frac{nm}{(1-\varpi)} = 0 \tag{11}$$

and

$$\frac{n\log\varpi}{1-\varpi^k} + \sum_{r=1}^{\infty} n_r \sum_{j=1}^{r} \frac{1}{(k+j-1)} = 0, \tag{12}$$

which, following Haldane (1941), is conveniently rewritten as

$$\frac{n\log\varpi}{1-\varpi^k} + \sum_{j=1}^{R} (k+j-1)^{-1} \sum_{i=j}^{R} n_i = 0, \tag{13}$$

where R is the highest observed value of r.

These equations are easily soluble by the usual maximum-likelihood iterative procedure. The components of the information matrix are the expected values of the quantities

$$-\frac{\partial^2 \log L}{\partial \varpi^2} = \frac{nk[1-(k+1)\varpi^k]}{\varpi^2(1-\varpi^k)^2} + \frac{nm}{(1-\varpi)^2},$$

$$-\frac{\partial^2 \log L}{\partial \varpi \partial k} = -\frac{n[1-(1-k\log \varpi)\varpi^k]}{\varpi(1-\varpi^k)^2},$$

$$-\frac{\partial^2 \log L}{\partial k^2} = \sum_{j=1}^{R} (k+j-1)^{-2} \sum_{r=j}^{R} n_r - \frac{n(\log \varpi)^2 \varpi^k}{(1-\varpi^k)^2}.$$

(14)

The iteration, however, is most conveniently carried out using the quantities (14) themselves, rather than their expected values. An example of the calculations for this method is given at the end of the next section.

<h2>8. AN ALTERNATIVE METHOD FOR THE SOLUTION OF THE MAXIMUM-LIKELIHOOD EQUATIONS</h2>

From equations (11) and (13) it follows that

$$\phi(\hat{\varpi}) = -\frac{\hat{\varpi} \log_e \hat{\varpi}}{1-\hat{\varpi}} = \frac{\hat{k}}{nm} \sum_{j=1}^{R} (\hat{k}+j-1)^{-1} \sum_{r=j}^{R} n_r,$$

(15)

where $\hat{\varpi}$ and $\hat{k}$ are maximum-likelihood estimates, and m is the sample mean. Equation (11) can be rewritten

$$\psi(\hat{\varpi}, \hat{k}) = \frac{\hat{\varpi}(1-\hat{\varpi}^k)m}{\hat{k}(1-\hat{\varpi})} = 1.$$

(16)

For a given value of k, equation (15) can be used to evaluate the corresponding ϕ and hence, from Table 4, ϖ; the value of ψ can then be calculated. This form of the equations was presented by David & Johnson (1952): they suggested an iterative method, equivalent in the present notation to

$$k_{i+1} = [k_i \psi(\varpi_i, k_i)],$$

and provided a table of values of $p = (1-\varpi)\varpi$ for values of ϕ (in their notation, y) $0.40 (0.05) 1.00$. This process, however, appears to converge rather slowly, and a solution can, in many cases, be reached more expeditiously by a trial-and-error process, starting with the moment, or some other inefficient, estimate of k. For this value of k the value of ψ can be evaluated, together with the slope of the curve ($\psi = \psi(k)$) defined by (15) and (16):

$$\psi'(k) = \frac{1}{\phi-\varpi}[\psi-m\varpi^{k+1}]\left[\frac{\phi}{k} - \frac{k}{nm}\sum_{j=1}^{R}(k+j-1)^{-2}\sum_{r=j}^{R} n_r\right] - \frac{\psi}{k}[1-\phi(\varpi^k)].$$

(17)

These values provide a second approximation to the root, which can then be found by a process of successive linear interpolation and extrapolation. Final adjustments may be made, if required, by the usual maximum-likelihood iterative procedure: this will usually be unnecessary, but the quantities (14) can be calculated, to provide an estimate of the asymptotic variance-covariance matrix of the estimates.

As with the moment method, extreme samples can occur for which the equations (15) and (16) have no solutions with $k > 0$. Unfortunately, the function $\psi(k)$ is of considerably more complicated a form than the function $\zeta(\varpi)$ occurring in the moment method, and the existence problem is correspondingly more difficult of solution. It can be shown that

$$\psi(0) = 1,$$

$$\psi'(0) = \frac{1}{n}\sum_{j=2}^{R}\frac{1}{(j-1)}\sum_{r=j}^{R} n_r + \frac{1}{2}\log \alpha,$$

where
$$\phi(\alpha) = 1/m,$$

and that
$$\lim_{k\to\alpha} \psi(k) = m(1-e^{-\theta})/\theta,$$

where
$$\theta = \frac{2}{nm} \sum_{j=2}^{R} (j-1) \sum_{r=j}^{R} n_r.$$

Clearly the conditions

$$\psi'(0) > 0, \quad \lim_{k\to\infty} \psi(k) < 1$$

are sufficient to ensure the existence of at least one solution, and I would conjecture that they are necessary, and also sufficient to ensure uniqueness, but have been unable to prove these results.

The question remains: What action is to be taken when the maximum-likelihood (or moment) method fails to give an acceptable solution? It seems reasonable to hope that one or other of the two limiting forms of the negative binomial, the Poisson distribution and Fisher's 'logarithmic series' distribution, will provide an adequate fit for all but some pathologically extreme (and highly improbable) samples which are unlikely to be fitted satisfactorily by any meaningful distribution.

9. An example of the maximum-likelihood calculations

The data are those already used in §4; details of the calculations are shown in Table 3. The quantity Σn_r, used in the calculation of ϕ, is tabulated in the third column of Table 3 (b), and the remainder of the calculations are shown in some detail in Table 3 (a). In fact, this table shows far more detail than need be recorded in practice; if the calculations are made on a desk machine only the first three columns of Table 3 (b) and the successive values of k, ϖ and ψ need be written down. (The weighted sums of quotients in the second column of Table 3 (a) can be accumulated on the machine, and they, together with the values of ϕ, $\log \varpi$ and ϖ^k, are used as soon as they are calculated, and need not be written down.) However, ϕ and ϖ^k should be recorded for the first trial value, for use in evaluating ψ', and if final adjustments are to be made by the iterative process, all the quantities tabulated in Table 3 (a) should be recorded for the last trial value of k. The fourth and fifth columns of Table 3 (b) are recorded for the purpose of calculating $\Sigma(k+j-1)^{-2} \sum_{r=j}^{R} n_r$; the fourth, $(0\cdot63+j-1)^2$, for use in calculating ψ'; the fifth, $(0\cdot493+j-1)^2$, for use in calculating the variance-covariance matrix. The only remaining values required in this section are

$$nm = 110, \quad m = 3\cdot4375.$$

As an initial trial value the moment estimate of k, $0\cdot633$, was rounded to two decimal places, giving
$$\psi(0\cdot63) = 0\cdot9922.$$

As a guide to the location of the second trial value, ψ' was evaluated for $k = 0\cdot63$. Accumulating on the machine,

$$\sum_{j=1} (0\cdot63+j-1)^{-2} \sum_{r=j}^{R} n_r = 92\cdot2942,$$

whence
$$\psi' = \frac{(0\cdot9922-0\cdot3179)(0\cdot7008-0\cdot5286)}{0\cdot2094} - \frac{0\cdot9922 \times 0\cdot3905}{0\cdot63}$$
$$= -0\cdot0605.$$

The second trial value was therefore taken as

$$0.63 - (1 - 0.9922)/0.0605 = 0.50.$$

The fourth value, $k = 0.493$, gives a value of ψ equal to 1 to four decimal places. There is no hope of improving further on this estimate by the trial and error method, without taking more decimal places, and using a more complicated interpolation formula, in the

Table 3. *Maximum-likelihood calculations for the example of § 8*

(a) Trial-and-error calculations

k	$\Sigma(k+j-1)^{-1} \sum\limits_{r=j}^{R} n_r$	ϕ	ϖ	$\log_e \varpi$	ϖ^λ	ψ	ψ'
0·63	77·0905	0·4415	0·2321	-1.46059	0·3984	0·9922	-0.0605
0·50	91·9239	0·4178	0·2124	-1.54930	0·4609	0·99952	
0·491	93·2185	0·4161	0·2110	-1.55590	0·4658	1·00016	
0·493	92·9270	0·4165	0·2113	-1.55450	0·4647	0·99996	

(b) Subsidiary tabulations

j	n_j	$\sum\limits_{r=j}^{R} n_r$	$(j-0.37)^2$	$(j-0.507)^2$	Expected values of n_j			
					Truncated Poisson	Truncated negative binomial		Log series
						Moments	M.L.	
1	11	32	0·3969	0·243049	4·01	10·31	10·80	13·32
2	6	21	2·6569	2·229049	6·64	6·44	6·36	5·85
3	4	15	6·9169	6·215049	7·33	4·33	4·17	3·43
4	5	11	13·1769	12·201049	6·07	3·01	2·87	2·26
5	0	6	21·4369	20·187049	4·02	2·13	2·03	1·59
6	1	6	31·6969	30·173049	2·22	1·53	1·47	1·16
7	0	5	43·9569	42·159049	1·05	1·11	1·07	0·88
8	2	5	58·2169	56·145049	0·43	0·81	0·79	0·67
9	1	3	74·4769	72·131049	0·16	0·60	0·59	0·53
10	0	2	92·7369	90·117049	0·05	0·44	0·44	0·42
11	1	2	112·9969	110·103049				
12	0	1	135·2569	132·089049	0·02	1·29	1·41	1·89
13	1	1	159·5169	156·075049				

table of ϕ. There is, in fact, no great need of any further improvement, but, if desired, final increments to k and ϖ can be calculated using the variance-covariance matrix, which requires calculation in any case.

The fifth column of Table 3 (b) shows values of

$$(0.493 + j - 1)^2,$$

and from these values (accumulating quotients on the machine)

$$\Sigma(0.493 + j - 1)^{-2} \sum_{r=j}^{R} n_r = 145.196464173.$$

Then, from this value and the entries in the last line of Table 3 (a), the estimate of the variance-covariance matrix, calculated from formulae (14), is

$$[V] = \begin{bmatrix} 554 \cdot 4186 & -94 \cdot 69298 \\ -94 \cdot 69298 & 19 \cdot 792976 \end{bmatrix}^{-1} = \begin{bmatrix} 0 \cdot 00986279 & 0 \cdot 0471853 \\ 0 \cdot 0471853 & 0 \cdot 276265 \end{bmatrix}.$$

Inserting the values $k = 0 \cdot 493$, $\varpi = 0 \cdot 2113$ into the left-hand sides of the maximum-likelihood equations (11) and (12), we obtain $0 \cdot 006$ and $-0 \cdot 0003$ respectively. Thus the final increments are

$$\begin{bmatrix} \delta\varpi \\ \delta k \end{bmatrix} = [V] \begin{bmatrix} 0 \cdot 006 \\ -0 \cdot 0003 \end{bmatrix} = \begin{bmatrix} 0 \cdot 000045 \\ 0 \cdot 00020 \end{bmatrix}.$$

Table 4. *The function* $\phi(x) = \dfrac{-x \log_e x}{1-x}$

x	0·00	0·01	0·02	0·03	0·04	0·05	0·06	0·07	0·08	0·09
0·0	0	0·0465	0·0798	0·1085	0·1341	0·1577	0·1796	0·2002	0·2196	0·2381
0·1	0·2558	0·2728	0·2891	0·3049	0·3201	0·3348	0·3491	0·3629	0·3764	0·3896
0·2	0·4024	0·4149	0·4271	0·4390	0·4507	0·4621	0·4733	0·4843	0·4950	0·5056
0·3	0·5160	0·5262	0·5362	0·5461	0·5558	0·5653	0·5747	0·5839	0·5930	0·6020
0·4	0·6109	0·6196	0·6282	0·6367	0·6451	0·6533	0·6615	0·6695	0·6775	0·6854
0·5	0·6931	0·7008	0·7084	0·7159	0·7233	0·7307	0·7380	0·7451	0·7522	0·7593
0·6	0·7662	0·7731	0·7800	0·7867	0·7934	0·8000	0·8066	0·8131	0·8195	0·8259
0·7	0·8322	0·8385	0·8447	0·8509	0·8570	0·8630	0·8690	0·8750	0·8809	0·8868
0·8	0·8926	0·8983	0·9041	0·9097	0·9154	0·9209	0·9265	0·9320	0·9374	0·9429
0·9	0·9482	0·9536	0·9589	0·9642	0·9694	0·9746	0·9797	0·9848	0·9899	0·9950

The maximum-likelihood estimates are therefore

$$\hat{\varpi} = 0 \cdot 2113 \pm 0 \cdot 0993, \quad \hat{k} = 0 \cdot 493 \pm 0 \cdot 526.$$

The variance-covariance matrix for the moment estimates, recalculated in terms of the maximum-likelihood estimates to provide a more valid comparison, is

$$\begin{bmatrix} 0 \cdot 01364125 & 0 \cdot 0693156 \\ 0 \cdot 0693156 & 0 \cdot 405868 \end{bmatrix}.$$

The ratio of the determinants of the two matrices, converted to a percentage efficiency, is

$$100 \times 0 \cdot 000498/0 \cdot 000732 = 68 \cdot 1 \%.$$

The efficiency is much the same for the two parameters, the actual values being $72 \cdot 3 \%$ for ϖ, and $68 \cdot 1 \%$ for k.

The maximum-likelihood estimate of k does not differ significantly from 0, which suggests that the data might be adequately fitted by Fisher's logarithmic series distribution. This, in fact, proves to be the case. The last four columns of Table 3 (b) show the expected numbers of cells with j breaks on the basis of four fitted distributions, a truncated Poisson (fitted by maximum likelihood), the two negative binomials fitted above, and the logarithmic series distribution (fitted by maximum likelihood). The fit of the Poisson distribution is obviously very poor; equally obviously the other three are all very good. The negative binomials

appear, superficially, to give a slightly better representation of the data, but this impression is largely due to the very good agreement in the first group. In fact the justification for using the negative binomial rather than the logarithmic series distribution comes, not from the data presented here, but from the whole series of experiments, only one of which is used here. In this series some distributions could be fitted adequately by the truncated Poisson, some by the logarithmic series distribution, and some not very satisfactorily by either, but the negative binomial form, with appropriate k, gave a good fit in all cases.

I am indebted to Dr C. E. Ford, of the Atomic Energy Research Establishment, Harwell, for permission to use his data in my examples, and to Miss A. D. Outhwaite, for drawing my attention to an omission from § 3.

REFERENCES

ANSCOMBE, F. J. (1950). Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, **37**, 358.

BLISS, C. I. & FISHER, R. A. (1953). Fitting the negative binomial distribution to biological data, with a note on the efficient fitting of the negative binomial. *Biometrics*, **9**, 176.

DAVID, F. N. & JOHNSON, N. L. (1952). The truncated Poisson. *Biometrics*, **8**, 275.

FISHER, R. A. (1941). The negative binomial distribution. *Ann. Eugen., Lond.*, **11**, 182.

GREENWOOD, M. & YULE, G. U. (1920). An enquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *J. R. Statist. Soc.* **83**, 255.

HALDANE, J. B. S. (1941). The fitting of binomial distributions. *Ann. Eugen., Lond.*, **11**, 179.

THE RANDOMIZATION ANALYSIS OF A GENERALIZED RANDOMIZED BLOCK DESIGN†

By M. B. WILK

Iowa State College

1. INTRODUCTION

(1·1) *The experimental situation and design*

Suppose that t treatments are given whose properties (yields, responses, effects, etc.) we wish to compare when they interact with a given set of rs experimental units, the latter being classified into r blocks, each containing $s = pt$ units. Suppose, further, that an experiment is carried out in which the treatments are applied at random to the experimental units, with the restriction that each treatment appears with p units in each of the r blocks. We refer to this design as the generalized randomized block design and note that it includes as special cases the completely randomized design ($r = 1, p > 1$) and the randomized block design ($r > 1, p = 1$).

The object of this paper is to study the basis for statistical inference which is provided by the randomization procedure.

(1·2) *Some previous work*

The introduction of the device of randomization in the statistical design of experiments is due to R. A. Fisher (1926, 1935). A brief review of some important contributions to the problem of inference from randomized experiments is given by Wilk (1953).

In particular, we note here that the pattern of the present study leans heavily on the finite model analyses given by Kempthorne (1952*a, b*). The means and variances of the analysis of variance sums of squares obtained by Welch (1937) and Pitman (1937) for the randomized block design, and by Kempthorne (1952*b*) for the completely randomized design, and the expectations under randomization of the analysis of variance mean squares for randomized blocks with non-additivity given by Kempthorne (1952*a*), derive as special cases from the results (Wilk, 1953) on which the present paper is based.

(1·3) *Experimental error and randomization*

In many experimental situations it seems reasonable to distinguish two sources of experimental error, namely, the failure of different experimental units treated alike to respond identically, and the inability to reproduce an applied treatment exactly. The first of these, which we shall refer to as the unit error, stems from variation among the experimental units. The second type, which we shall call the technical error, stems from limitations on experimental technique.

Generally, the unit error is to be regarded as a fixed quantity associated with any given experimental unit; while the technical error may often be idealized as a random variable, say following a normal distribution with mean 0. The process of randomization is, usually, irrelevant to our conception of the technical errors, but is of critical importance in our treatment of the unit errors. It might in fact be said that the main function of randomization in experimental design is to control, in a statistical sense, the unit errors.

To focus attention on the basis for statistical inference which is provided by the randomization procedure we shall in this paper assume that the only important source of experimental error is the unit errors.

2. THE ANALYSIS OF VARIANCE

(2·1) The conceptual underlying population

With the possible application of each of the treatments to each of the experimental units we associate a real (unknown) number. This defines a set of rst numbers which we take to be the conceptual underlying population for this experiment. Since the fact of the experimental situation is that each unit can be 'used' only once, the population defined is conceptual in the sense that only a subset of rs of the numbers of interest can be observed. The scope of a statistical inference for this situation can be delineated by noting that the conceivable totality of experimental information would be given by applying each treatment to every experimental unit and observing the response.

(2·2) The population model

Let $i = 1, 2, ..., r$ denote the block number.

Let $j = 1, 2, ..., s$ denote the experimental unit number within each block, where $s = pt$.

Let $k = 1, 2, ..., t$ denote the treatment number.

Let y_{ijk} represent the (conceptual) response which would be obtained if treatment k were applied to the jth unit in the ith block. Thus our underlying population is the set of (conceptual) unknown numbers $\{y_{ijk}\}$.

We will employ the usual dot convention for means, for example, $y_{..k} = \sum_{ij} y_{ijk}/(rs)$.

We now define
$$\mu = y_{...},$$
$$b_i = y_{i..} - y_{...},$$
$$t_k = y_{..k} - y_{...},$$
$$(bt)_{ik} = (y_{i.k} - y_{i..}) - (y_{..k} - y_{...}),$$
$$e_{ij} = y_{ij.} - y_{i..},$$
$$n_{ijk} = (y_{ijk} - y_{ij.}) - (y_{i.k} - y_{i..}),$$

where $i = 1, 2, ..., r$; $j = 1, 2, ..., s$; $k = 1, 2, ..., t$.

These quantities may be given a physical interpretation.

μ is the (conceptual) overall mean yield which would be obtained if each treatment were applied to every unit in every block.

b_i is the difference between the (conceptual) mean yield of all treatments on all units of block i and μ, and may be thought of as the effect attributable to the ith block.

In an analogous way t_k may be thought of as the effect attributable to the kth treatment. We note that by definition this is the average effect over all experimental units.

$(bt)_{ik}$ is the difference between the effect of treatment k on block i and t_k. Thus it is a measure of the extent to which treatment k and block i interact and will be called the block-treatment interaction.

e_{ij} gives the difference between the (conceptual) mean of the yields of all treatments on unit j of block i and the mean over the whole block. It, therefore, measures the extent to which the jth unit deviates from the other units of block i and will be called the unit error.

The set $\{e_{ij}\}, j = 1, 2, \ldots, s$, can be used to give a measure of the heterogeneity of units within the ith block.

n_{ijk} is the difference between the effect of treatment k on unit j of block i and the effect of treatment k over all of block i. It is, therefore, a measure of the extent to which treatment k and unit j of block i interact and will be called the unit-treatment interaction.

The following equation is an algebraic indentity:

$$y_{ijk} = \mu + b_i + t_k + (bt)_{ik} + e_{ij} + n_{ijk}.$$

From the definition of the quantities it follows that

$$\sum_i b_i = \sum_k t_k = \sum_i (bt)_{ik} = \sum_k (bt)_{ik} = \sum_j e_{ij} = \sum_j n_{ijk} = \sum_k n_{ijk} = 0.$$

Algebraically the sets of numbers $\{b_i\}$, $\{t_k\}$, $\{(bt)_{ik}\}$, $\{e_{ij}\}$ and $\{n_{ijk}\}$ are pairwise independent in the sense that, for example, $t_k = 0$ for all k does not imply $(bt)_{ik} = 0$ for all k. In particular, if $(bt)_{ik} = 0$ for all i and k, then algebraically this does not imply that $n_{ijk} = 0$ for all i and k. On the other hand, the physical situation is such that if all the block-treatment interactions were zero, then one would expect that all the unit-treatment interactions would be zero. This follows from the fact that the experimental units are blocked so as to be more homogeneous within a block than from block to block. If so, then the lack of a differential effect from block to block would lead one to expect that the differential effect from unit to unit within a block would be negligible.

The converse is, of course, not true. If the units within a block are sufficiently homogeneous then the unit-treatment interactions will be small in absolute value. But this does not preclude important differences between blocks and hence the possible existence of non-additive effects from block to block, i.e. block-treatment interactions. It would appear therefore that if the blocking of experimental units is successful, then the n_{ijk} will be negligible while the $(bt)_{ik}$ may be important.

The essential point here is that it may not be unrealistic to assume that the treatments react additively (i.e. unit-treatment interactions are zero) within a block even though they react non-additively from block to block.

In many instances, whether we assume the unit-treatment interactions are zero or not, it may be reasonable to assume that the variability of units within a block is essentially the same for all blocks. This assumption could be idealized as

$$\sum_j e_{ij}^2 = (s-1)\sigma^2;$$

$$\sum_j n_{ijk}^2 = (s-1)\sigma_n^2 \quad (i = 1, 2, \ldots, r; \; k = 1, 2, \ldots, t).$$

The importance of this discussion is in indicating the direction of simplifying assumptions in the analysis of the design.

(2·3) *The statistical model*

In actually carrying out the experiment, we will in fact observe only a (restricted) random sample of size rs from the set of rst numbers $\{y_{ijk}\}$. Let x_{ikf} denote the observation obtained from the fth replication of the kth treatment in the ith block, where $i = 1, 2, \ldots, r$; $k = 1, 2, \ldots, t$; and $f = 1, 2, \ldots, p$ for each (ik). Thus $\sum_f x_{ikf}$ represents the observed total response from all units in block i to which treatment k has been applied.

To write an explicit model for $\sum_f x_{ikf}$ we now define some additional quantities. Let†

$$D_{ij}^k = 1 \quad \text{if treatment } k \text{ falls on unit } j \text{ of block } i$$
$$= 0 \quad \text{otherwise.}$$

Because random methods of allocation are employed, the D_{ij}^k may be treated as random variables, and from the design of the experiment it is easy to specify certain of the distributional properties of the D_{ij}^k. For example,

$$P(D_{ij}^k = 1) = P \text{ (treatment } k \text{ falls on unit } j \text{ of block } i)$$
$$= p/s.$$

Hence

$$E(D_{ij}^k) = p/s,$$

where we follow usual convention and define $E(x)$ to be the mathematical expectation of x. More detail on the D_{ij}^k is given by Wilk (1953).

It is easy to see that

$$\sum_f x_{ikf} = p[\mu + b_i + t_k + (bt)_{ik}] + \sum_j e_{ij} D_{ij}^k + \sum_j n_{ijk} D_{ij}^k.$$

This relation we shall call the statistical model. This formulation exhibits explicitly just what are the random variables involved in the model, namely, the D_{ij}^k which take on the values 0 or 1 with known probabilities. We note that it is the physical act of randomization in allocation of treatments to units that permits us to treat the D_{ij}^k as random variables, and hence provides some basis for statistical inference.

(2·4) The analysis of variance table

The primitive analysis of variance for this design is simply a breakdown of the sum of squares of deviations of individual observations from their mean into additive components which can be attributed to various sources. The analysis of variance has proved useful in the statistical analysis of experiments in the estimation of components of variation, in the estimation of the variance of estimates of treatment comparisons, and in making tests of significance.

As before, we use the dot convention for means, e.g. $x_{i..} = \sum_{kf} x_{ikf}/s$. The algebraic detail of the analysis of variance is given in Table 1.

We can now employ the statistical model for $\sum_f x_{ikf}$, and certain properties of the D_{ij}^k, to derive the expectations under randomization of the analysis of variance mean squares. The detailed algebra is given by Wilk (1953).

The results are tabulated in Table 2.

It is of interest to note that if the n_{ijk} are not all zero, then even if $t_k = 0$ for all k, the expectation of the treatment mean square is not equal to the expectation of the error mean square. Similarly for the interaction mean square. If the n_{ijk} are small compared with the e_{ij}, or if t is large, the bias is negligible. The assumptions

$$\sum_j e_{ij}^2 = (s-1)\sigma^2, \quad \text{all } i,$$

$$\sum_j n_{ijk}^2 = (s-1)\sigma_n^2, \quad \text{all } i \text{ and } k,$$

do not affect the above discussion.

† A more formal definition of the D_{ij}^k, in which they appear as characteristic set functions of set-valued random variables, is given by Wilk (1953).

Clearly, for small t, a meaningful comparison of the treatment mean square (or interaction mean square) with the error mean square depends heavily on the assumption that the n_{ijk} are negligible.

Table 1. *Analysis of variance*

Due to	Degrees of freedom	Sum of squares	Mean squares
Blocks	$(r-1)$	$B = s \sum_i (x_{i..} - x_{...})^2$	$B^* = \dfrac{B}{r-1}$
Treatments	$(t-1)$	$T = rp \sum_k (x_{.k.} - x_{...})^2$	$T^* = \dfrac{T}{(t-1)}$
Interactions	$(r-1)(t-1)$	$I = p \sum_{ik} (x_{ik.} - x_{i..} - x_{.k.} + x_{...})^2$	$I^* = \dfrac{I}{(r-1)(t-1)}$
Error	$rt(p-1)$	$R = \sum_{ikj} (x_{ikj} - x_{ik.})^2$	$R^* = \dfrac{R}{rt(p-1)} = v$
Total	$rs-1$	$G = \sum_{ikj} (x_{ikj} - x_{...})^2$	

Table 2. *Expectations of mean squares under randomization*

Due to	Mean square	Expectation of mean square
Blocks	B^*	$\dfrac{1}{rt(s-1)} \sum_{ijk} n_{ijk}^2 + \dfrac{pt}{(r-1)} \sum_i b_i^2$
Treatments	T^*	$\dfrac{1}{r(s-1)} \sum_{ij} e_{ij}^2 + \dfrac{(t-2)}{rt(s-1)(t-1)} \sum_{ijk} n_{ijk}^2 + \dfrac{rp}{(t-1)} \sum_{ik} (bt)_{ik}^2$
Interactions	I^*	$\dfrac{1}{r(s-1)} \sum_{ij} e_{ij}^2 + \dfrac{(t-2)}{rt(s-1)(t-1)} \sum_{ijk} n_{ijk}^2 + \dfrac{p}{(r-1)(t-1)} \sum_{ik} (bt)_{ik}^2$
Error	R^*	$\dfrac{1}{r(s-1)} \sum_{ij} e_{ij}^2 + \dfrac{1}{rt(s-1)} \sum_{ijk} n_{ijk}^2$

In contrast with similar results based on normal theory, the expectation of the mean square for blocks does not contain all components which appear in the expectation of the error mean square. The same remark applies to a comparison of blocks mean square with interaction mean square when all $(bt)_{ik} = 0$. It is apparent that an analysis of variance test of significance of block effects, for the situation under consideration, cannot be justified by randomization.

(2·5) *Some randomization moments*

In Table 3 we give some moments under randomization of the analysis of variance sums of squares, under the hypothesis that $t_k = (bt)_{ik} = 0$ for all i and k, and using the simplifying assumptions:

$$n_{ijk} = 0, \qquad \text{for all } i, j, k,$$
$$\sum_j e_{ij}^2 = (s-1)\sigma^2, \qquad \text{for all } i,$$
$$\sum_j e_{ij}^4 = (s-1)(2s-3)\sigma^4/s, \qquad \text{for all } i.$$

We use G_0, B_0, T_0, I_0 and R_0 to denote the values of G, B, T, I and R, respectively, under the indicated conditions.

The results of Table 3 are derived from more general expressions, given by Wilk (1953), which use no homogeneity assumptions regarding $\sum_j e_{ij}^2$ and $\sum_j e_{ij}^4$. The algebraic development is also detailed in the same paper.

Table 3. *Some randomization moments under simplifying assumptions*

$$E(G_0) = G_0 = pt\sum_i b_i^2 + rt(s-1)\sigma^2 \qquad V(G_0) = 0$$

$$E(B_0) = B_0 = pt\sum_i b_i^2 \qquad V(B_0) = 0$$

$$E(T_0) = (t-1)\sigma^2 \qquad V(T_0) = 2(t-1)\sigma^4\left(1-\frac{1}{rp}\right)$$

$$E(I_0) = (r-1)(t-1)\sigma^2 \qquad V(I_0) = 2(r-1)(t-1)\sigma^4\left(1-\frac{r-1}{rp}\right)$$

$$E(R_0) = rt(p-1)\sigma^2 \qquad V(R_0) = \frac{2r(t-1)(p-1)}{p}\sigma^4$$

$$\operatorname{cov}(T_0, I_0) = -\frac{2(t-1)(r-1)}{rp}\sigma^4 \qquad \operatorname{cov}(T_0, R_0) = -\frac{2(p-1)(t-1)}{p}\sigma^4$$

$$\operatorname{cov}(I_0, R_0) = -\frac{2(p-1)(t-1)(r-1)}{p}\sigma^4$$

It is of some interest to note that for large r or p, $V(T_0)$ approaches $2(t-1)\sigma^4$; for large p, $V(I_0)$ approaches $2(r-1)(t-1)\sigma^4$; for large p, $V(R_0)$ becomes independent of p and approaches $2r(t-1)\sigma^4$; if $p = 1$, $V(I_0)$ becomes $(2r-1)(t-1)\sigma^4/r$, and thence for large r becomes independent of r. The correspondence to and divergence from the corresponding moments based on normal theory will be apparent.

The correlation between T_0 and I_0 is $-\sqrt{\dfrac{(r-1)}{(rp-1)(rp-r+1)}}$, which is -1 for $p = 1$, and goes to 0 as p increases. The correlation between T_0 and R_0 is $-\sqrt{\dfrac{p-1}{rp-1}}$, which for large p goes to $-\sqrt{\dfrac{1}{r}}$, and for large r goes to zero. The correlation between I_0 and R_0 is

$$-\sqrt{\frac{(r-1)(p-1)}{(rp-r-1)}},$$

which for large p goes to $-\sqrt{\dfrac{(r-1)}{r}}$, and for large r goes to -1.

3. ESTIMATION

This section deals with the estimation of certain functions of the parameters of the population model. Variances of the estimates under various conditions are given, and the estimation of these variances is considered. Some additional results, as well as the algebraic detail, are given by Wilk (1953).

The following general notation is used:

$V(x)$ is the variance of x under no assumptions.

$V_1(x)$ is the variance of x under the homogeneity assumptions that

$$\sum_j e_{ij}^2 = (s-1)\sigma^2, \quad \sum_j n_{ijk}^2 = (s-1)\sigma_n^2 \quad \text{and} \quad \sum_j e_{ij} n_{ijk} = 0,$$

for each $i = 1, 2, ..., r$ and $k = 1, 2, ..., t$.

$V^0(x)$ is the variance of x under the assumption that all the n_{ijk} are zero.

v denotes the analysis of variance error mean square.

(a) An unbiased estimate of μ is $\hat{\mu} = x_{...}$:

$$V(\hat{\mu}) = \frac{1}{r^2 t s(s-1)} \sum_{ijk} n_{ijk}^2, \text{ and there appears to be no reasonable way to estimate } \sum_{ijk} n_{ijk}^2.$$

$V^0(\hat{\mu}) = 0$, so μ is known without error if the n_{ijk} are all zero.

(b) An unbiased estimate of b_t is $\hat{b}_t = x_{t..} - x_{...}$:

$$V(\hat{b}_i) = \frac{(r-2)}{rts(s-1)} \sum_{jk} n_{ijk}^2 + \frac{1}{r^2 t s(s-1)} \sum_{ijk} n_{ijk}^2.$$

$V_1(\hat{b}_i) = \frac{r-1}{rs} \sigma_n^2$. No reasonable estimate of either $V(\hat{b}_i)$ or $V_1(\hat{b}_i)$ is available.

$V^0(\hat{b}_j) = 0$, so the block effects are known precisely if all the n_{ijk} are zero.

(c) An unbiased estimate of t_k is $\hat{t}_k = x_{.k.} - x_{...}$:

$$V(\hat{t}_k) = \frac{1}{r^2 s(s-1)} \left[(t-1) \sum_{ij} e_{ij}^2 + 2(t-2) \sum_{ij} e_{ij} n_{ijk} + (t-3) \sum_{ij} n_{ijk}^2 + \frac{1}{t} \sum_{ijk} n_{ijk}^2 \right].$$

$V_1(\hat{t}_k) = \frac{(t-1)}{rs}(\sigma^2 + \sigma_n^2) - \frac{1}{rs}\sigma_n^2$, and $\frac{(t-1)}{rs} v$ tends to overestimate $V_1(\hat{t}_k)$.

$V^0(\hat{t}_k) = \frac{(t-1)}{r^2 s(s-1)} \sum_{ij} e_{ij}^2$, and $\frac{(t-1)}{rs} v$ is an unbiased estimate of $V^0(\hat{t}_k)$.

(d) An unbiased estimate of $(bt)_{ik}$ is $\widehat{(bt)}_{ik} = x_{ik.} - x_{i..} - x_{.k.} + x_{...}$:

$$V(\widehat{bt})_{ik} = \frac{1}{rs(s-1)} \left[(t-1)(r-2) \sum_j e_{ij}^2 + \frac{(t-1)}{r} \sum_{ij} e_{ij}^2 \right.$$
$$+ 2(t-2)(r-2) \sum_j e_{ij} n_{ijk} + \frac{2(rt-r-2t+1)}{r} \sum_{ij} e_{ij} n_{ijk}$$
$$\left. + (rt - 3r - 2t + 4) \sum_j n_{ijk}^2 + \frac{(t-3)}{r} \sum_{ij} n_{ijk}^2 \right].$$

$V_1(\widehat{bt})_{ik} = \frac{(r-1)(t-1)}{rs}(\sigma^2 + \sigma_n^2) - \frac{r-1}{rs}\sigma_n^2$, and so $\frac{(r-1)(t-1)}{rs} v$ tends to overestimate $V_1(\widehat{bt})_{ik}$.

(e) An unbiased estimate of the treatment mean $(\mu + t_k)$ is $(\hat{\mu} + \hat{t}_k) = x_{.k.}$:

$$V(\hat{\mu} + \hat{t}_k) = \frac{(t-1)}{r^2 s(s-1)} \sum_{ij} (e_{ij}^2 + 2 e_{ij} n_{ijk} + n_{ijk}^2).$$

$V_1(\hat{\mu} + \hat{t}_k) = \frac{(t-1)}{rs}(\sigma^2 + \sigma_n^2)$, and $\frac{(t-1)}{rs} v$ is an unbiased estimate of $V_1(\hat{\mu} + \hat{t}_k)$.

$V^0(\hat{\mu} + \hat{t}_k) = \frac{(t-1)}{r^2 s(s-1)} \sum_{ij} e_{ij}^2$, and $\frac{(t-1)}{rs} v$ is an unbiased estimate of $V^0(\hat{\mu} + \hat{t}_k)$.

(f) An unbiased estimate of the treatment contrast $\sum_k \rho_k t_k$, with $\sum_k \rho_k = 0$, is

$$\sum_k \rho_k \hat{t}_k = \sum_k \rho_k (x_{.k.} - x_{...});$$

$$V(\sum_k \rho_k \hat{t}_k) = \frac{1}{r^2 p(s-1)} \left[(\sum_k \rho_k^2) \sum_{ij} e_{ij}^2 + 2 \sum_{ijk} \rho_k^2 e_{ij} n_{ijk} + \sum_{ijk} \rho_k^2 n_{ijk}^2 - \frac{1}{t} \sum_k (\sum_{ij} \rho_k n_{ijk})^2 \right].$$

$$V_1(\sum_k \rho_k \hat{t}_k) = \frac{1}{rp} (\sum_k \rho_k^2)(\sigma^2 + \sigma_n^2) - \frac{1}{r^2 s(s-1)} \sum_{ij} (\sum_k \rho_k n_{ijk})^2, \text{ and so } \frac{1}{rp} (\sum_k \rho_k^2) r \text{ tends to}$$

overestimate $V_1(\sum_k \rho_k \hat{t}_k)$.

$$V^0(\sum_k \rho_k \hat{t}_k) = \frac{1}{r^2 p(s-1)} (\sum_k \rho_k^2) \sum_{ij} e_{ij}^2, \text{ and this is estimated unbiasedly by } \frac{1}{rp} (\sum_k \rho_k^2) r.$$

We close this section with a brief discussion of special cases. If $r = 1$, $p > 1$ (completely randomized design), the discussion of block effects and interactions is to be ignored, and the remaining results carry over directly.

For the case of $p = 1$, $r > 1$ (randomized block design) the situation is somewhat different in that the analysis of variance error mean square which we have been using becomes non-existent. If we can assume that all block-treatment interactions are zero, then the inter-action mean square in the randomized block design has expectation

$$\frac{1}{r(t-1)} \sum_{ij} e_{ij}^2 + \frac{(t-2)}{rt(t-1)^2} \sum_{ijk} n_{ijk}^2,$$

and for most of the cases discussed above, the variances of estimates may be estimated unbiasedly.

Kempthorne (1952 a, b) has discussed in detail the randomization analysis of the randomized block and completely randomized designs.

4. Tests of significance

In this section we consider tests of significance of a number of null hypotheses (hypotheses of equivalence) of possible interest. The object of the test is to obtain a measure of the adequacy of the experiment to indicate conclusions.

To employ a randomization test of a null hypothesis we require no assumptions about the form of the frequency distribution of the observations. A real-valued function of the observations is selected which will reflect (in a monotone increasing fashion, say) the deviation from equivalence of the treatments.

Having obtained a set of numbers for the particular experimental arrangement, we associate these numbers with the experimental units from which they derived. We then evaluate the function selected for the actual disposition of the treatments and for all other dispositions possible under the restriction of the design. A measure of the strength of the evidence against the null hypothesis (i.e. the level of significance† of the experiment) is the proportion of the values of the function which exceed the 'observed' value. The initial introduction of randomization enables a probabilistic interpretation.

In general, the amount of computation implied by the above procedure is prohibitive. Consequently, several writers (Welch, 1937; Pitman, 1937; Kempthorne, 1952 a, b) have studied the possible approximation to the randomization test of certain procedures which

† A high level of significance corresponds to a low proportion.

derive from normal theory assumptions. Since the normal theory criteria fulfil the requirement for a randomization test function, the approach has been to examine the correspondence by comparing randomization means and variances of these criteria with their normal theory analogues. We proceed to make a similar examination for several null hypotheses of possible interest in the generalized randomized block design. The normal theory analysis for this design has been outlined by Wilk (1953).

In all that follows in this section we make the assumption that the n_{ijk} are all zero, and that $\sum_j e_{ij}^2 = (s-1)\sigma^2$, $\sum_j e_{ij}^4 = (s-1)(2s-3)\sigma^4/s$ $(i = 1, 2, \ldots, r)$.

(a) Consider the null hypothesis that the treatments all react identically on every experimental unit. This implies that $y_{ijk} = y_{ij}$, for all i, j, k, and hence $t_k = (bt)_{ijk} = n_{ijk} = 0$ for all i, j, k.

The normal theory test of this hypothesis may be based on the criterion

$$(T_0 + I_0)/(T_0 + I_0 + R_0),$$

which under the usual normal theory assumptions and under the present null hypothesis follows a Beta distribution

$$f(q) = K q^{\frac{1}{2}f_1 - 1}(1-q)^{\frac{1}{2}f_2 - 1} \quad (0 < q < 1),$$

where $f_1 = r(t-1)$, $f_2 = rt(p-1)$.

Let U and U^* denote the criterion above under normal theory and randomization respectively. Then $0 < U < 1$ and $0 < U^* < 1$. Since the denominator of U^* is constant under randomization $E(U^*) = \dfrac{(t-1)}{(s-1)}$ and $V(U^*) = \dfrac{2(t-1)(p-1)}{rp(s-1)^2}$. Under normal theory,

$E(U) = \dfrac{(t-1)}{(s-1)}$, $V(U) = \dfrac{2t(t-1)(p-1)}{(rs-r+2)(s-1)^2}$. Thus $E(U^*) = E(U)$; and for large s or small r, $V(U)$ approaches $V(U^*)$.

If we can judge the correspondence of the distributions of U and U^* from the comparison of their ranges, means and variances, it would appear that the randomization test may be approximated by the usual test based on normal theory.

To better the approximation to the randomization distribution of U^* by a Beta distribution, we might use a Beta variate Q, with density

$$f(Q) = K Q^{\frac{1}{2}f_1 - 1}(1-Q)^{\frac{1}{2}f_2 - 1} \quad (0 < Q < 1),$$

such that

$$E(Q) = \frac{f_1}{f_1 + f_2} = \frac{(t-1)}{(s-1)} = E(U^*),$$

and

$$V(Q) = \frac{2 f_1 f_2}{(f_1 + f_2)^2 (f_1 + f_2 + 2)} = \frac{2(t-1)(p-1)}{rp(s-1)^2} = V(U^*).$$

Then

$$f_1 = \frac{(t-1)(rs-2)}{(s-1)}, \quad f_2 = \frac{t(p-1)(rs-2)}{(s-1)}.$$

For different assumptions regarding the quantities $\sum_j e_{ij}^2$ and $\sum_j e_{ij}^4$ $(i = 1, 2, \ldots, r)$, we would thus arrive at different Beta distributions to approximate the randomization test. Welch (1937) considered such an adjustment for the case of the randomized block design when the blocks do not exhibit equal variation.

(b) We consider now a situation in which we can assume that $(bt)_{jk} = 0$ for all i and k, and wish to test the null hypothesis that $t_k = 0, k = 1, 2, \ldots, t$.

The criterion suggested by normal theory for this situation is $T_0 (T_0 + I_0 + R_0)$. Let W denote this quantity under normal theory, and W^* denote it under randomization.

Then under the null hypothesis, $0 < W < 1$, $0 < W^* < 1$, and W is a Beta variate with

$$E(W) = \frac{(t-1)}{r(s-1)} \text{ and } V(W) = \frac{2(t-1)(rs-r-t-1)}{r^2(s-1)^2(rs-r-2)}; \text{ while } E(W^*) = \frac{(t-1)}{r(s-1)} \text{ and}$$

$$V(W^*) = \frac{2(t-1)}{r^2(s-1)^2}\left(1 - \frac{1}{rp}\right).$$

Thus $E(W) = E(W^*)$, and for large values of p or r $V(W)$ and $V(W^*)$ are approximately equal.

Of course as in (a) we can define a Beta variate which will have the same mean and variance as W^* and, perhaps, in this way better the approximation.

A special case of interest is that in which $p = 1$, $r > 1$, i.e. the randomized block design. In that case, R does not exist, and carrying out a meaningful analysis of variance test of significance for that design hinges upon the assumption that the block-treatment interactions are zero.

Also, of particular concern here is the case of $r = 1$, $p > 1$, i.e. the completely randomized design. In that case, I does not exist, and the test we have discussed is the one of interest for that design.

It is a pleasure to acknowledge the assistance of Prof. O. Kempthorne in the preparation of this paper.

REFERENCES

FISHER, R. A. (1926). The arrangement of field experiments. *J. Minist. Agric.* **33**, 503–13.
FISHER, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.
KEMPTHORNE, O. (1952a). *The Design and Analysis of Experiments*. New York: John Wiley and Sons.
KEMPTHORNE, O. (1952b). The randomization theory of experimental inference. Paper delivered to the Amer. Stat. Ass. and Biometrics Soc. Publication pending.
PITMAN, E. J. G. (1937). Significance tests which may be applied to samples from any population. III. The analysis of variance test. *Biometrika*, **29**, 322–35.
WELCH, B. L. (1937). On the z-test in randomized blocks and latin squares. *Biometrika*, **29**, 21–52.
WILK, M. B. (1953). The randomization analysis of a generalized randomized block design. M.S. thesis, Department of Statistics, Iowa State College.

SOME QUICK SIGN TESTS FOR TREND IN LOCATION AND DISPERSION

By D. R. COX

Statistical Laboratory, University of Cambridge

AND A. STUART

Division of Research Techniques, London School of Economics

1. INTRODUCTION AND SUMMARY

Many distribution-free tests have been devised to test the hypothesis of randomness of a series of N observations, i.e. the hypothesis that N independent random variables have the same continuous distribution function. Of these, the rank correlation tests are the most efficient tests against normal trend alternatives, but others are of some use in situations where speed and simplicity of computation are important.

In this paper, we discuss a class of simple sign tests, considered first as tests against trend in location. Optimum tests are found from the standpoint of asymptotic relative efficiency (a measure of local power in large samples), and it appears that the best of these tests may be preferred to the other simple tests considered in the literature, although they are, of course, less efficient than the rank correlation tests.

Similar tests are available for trend in dispersion, and the efficiency of these, in the normal situation, is investigated and compared with the test based on the maximum-likelihood estimator. Finally, we add a few remarks on sequential sign tests.

Readers not interested in the theory should look at §§ 10, 11 and 14, where there are brief statements of the tests and numerical examples.

2. THE SIGN TESTS FOR TREND IN LOCATION

We consider a series of N independent observations from a standardized normal regression model with an upward trend, i.e.

$$H_1: \quad y_i = \alpha + \Delta i + \epsilon_i \quad (i = 1, 2, ..., N),$$

where $\Delta \geqslant 0$ and the ϵ_i are independent standardized normal variates. We wish to test the null hypothesis

$$H_0: \quad \Delta = 0,$$

using a distribution-free test statistic, so that our test will remain valid whatever the continuous distribution of the ϵ terms in the model, although naturally its efficiency will vary with the form of the distribution.

The most efficient known distribution-free tests of H_0 are those based on the rank correlation coefficients (Stuart, 1954), but our object here is specifically to find tests which are quick and simple to compute. We define for $i < j$ the score

$$h_{ij} = \begin{cases} +1 & \text{if} \quad y_i > y_j, \\ 0 & \text{if} \quad y_i < y_j. \end{cases}$$

h_{ij} is thus based on a comparison of the ith and jth in the series of observations. The distribution of the observations will be assumed continuous so that the possibility of ties can be ignored (see, however, §16).

We confine ourselves throughout to comparisons of *independent* pairs of observations, i.e. no observation is compared with more than one other observation. (This is in contrast to the procedure used in calculating the rank correlation coefficients, where every observation is compared with every other observation in the series.) Since there are N observations, there can be no more than $\frac{1}{2}N$ such independent comparisons. We now assume N to be even and always take $i < j$. Our problem is to find the set of comparisons and the appropriate weights w_{ij} which will make the statistic

$$S = \Sigma w_{ij} h_{ij} \tag{1}$$

as efficient a test of H_0 as possible. The summation in (1) contains $\frac{1}{2}N$ terms, no suffix being repeated. All tests of the form (1) are distribution-free, since on the null hypothesis any h_{ij} is a $0-1$ variate with probabilities $(\frac{1}{2}, \frac{1}{2})$, whatever the distribution of the ϵ_i.

3. Asymptotic relative efficiency

We shall use as our criterion of efficiency the asymptotic relative efficiency (A.R.E.) of a test. If there are two consistent tests, s and t, of a hypothesis $H_0: \Delta = 0$, the A.R.E. is the reciprocal of the ratio of sample sizes required to attain the same power against the same alternative hypothesis H_1, taking the limit as the sample size N tends to infinity and as H_1 tends to H_0. (This second limiting process is necessary to keep the power of consistent tests bounded away from 1.) Pitman (1948) and Noether (1955) have shown that, if s and t both have normal limiting distributions on H_0 and H_1, the A.R.E. of s compared to t is given by

$$\text{A.R.E. } (s, t) = \lim_{N \to \infty} \left(\frac{R^2(s)}{R^2(t)} \right)^{1/r}, \tag{2}$$

where

$$R^2(X) = \left\{ \left[\frac{\partial}{\partial \Delta} E(X \mid \Delta) \right]_{\Delta=0} \right\}^2 \bigg/ D^2(X \mid \Delta = 0), \tag{3}$$

provided that r satisfies the equations

$$\lim_{N \to \infty} R^2(s) N^{-r} = R_1, \quad \lim_{N \to \infty} R^2(t) N^{-r} = R_2. \tag{4}$$

Here E and D^2 denote mean and variance as usual, while R_1 and R_2 are constants independent of N. The interpretation of the A.R.E. is discussed critically in §9 below.

4. The best sign test

Since h_{ij} is a 0–1 variate,

$$E(h_{ij}) = \text{prob} (y_i > y_j),$$

and as $(y_j - y_i)$ is a normal variate with mean $(j-i)\Delta$ and variance 2 this is

$$E(h_{ij}) = G\left\{ -\frac{(j-i)\Delta}{\sqrt{2}} \right\}, \tag{5}$$

where

$$G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}t^2} dt.$$

Now

$$\left[\frac{\partial}{\partial x} G(x) \right]_{x=0} = \frac{1}{\sqrt{(2\pi)}}, \tag{6}$$

so that, by (5) and (6),

$$E'(h_{ij}) = \left[\frac{\partial}{\partial \Delta} E(h_{ij})\right]_{\Delta=0} = \frac{(i-j)}{2\sqrt{\pi}}. \tag{7}$$

We now write $(j-i) = r_{ij}$. Using (7) in (1), we obtain

$$E'(S) = \Sigma w_{ij} E'(h_{ij}) = -\frac{1}{2\sqrt{\pi}} \Sigma w_{ij} r_{ij}. \tag{8}$$

We also have

$$D^2(S\,|\,\Delta=0) = \Sigma w_{ij}^2 V(h_{ij}\,|\,H_0) = \tfrac{1}{4}\Sigma w_{ij}^2. \tag{9}$$

Equations (3), (8) and (9) give

$$R^2(S) = \frac{1}{\pi} \frac{(\Sigma w_{ij} r_{ij})^2}{\Sigma w_{ij}^2}, \tag{10}$$

and we now wish to maximize (10) to obtain the highest possible A.R.E. We do this in two stages. First we maximize (10) with respect to the w_{ij}, regarding the r_{ij} as fixed, and then we choose the supremum of these maxima for variations in the r_{ij}.

To maximize (10) for fixed r_{ij} and variation in the w_{ij}, we must maximize $\Sigma w_{ij} r_{ij}$ subject to Σw_{ij}^2 being held constant, i.e. we must unconditionally maximize

$$F = \Sigma w_{ij} r_{ij} - \lambda \Sigma w_{ij}^2.$$

It is clear from the conditions of the problem that each w_{ij} will be a function of the corresponding r_{ij}, so that on differentiating F for a stationary value we get

$$r_{ij} + w_{ij} \frac{\partial r_{ij}}{\partial w_{ij}} - 2\lambda w_{ij} = 0,$$

i.e.

$$\frac{r_{ij}}{w_{ij}} + \frac{\partial r_{ij}}{\partial w_{ij}} = 2\lambda.$$

This is satisfied by

$$w_{ij} = \lambda r_{ij}, \tag{11}$$

so that the required set of weights are proportional to the distances apart of the observations compared. The stationary value is a maximum. Substituting (11) into (10), we have

$$R^2(S) = \frac{1}{\pi} \Sigma r_{ij}^2. \tag{12}$$

This is the maximum value of $R^2(S)$ for a fixed set of r_{ij}. The r_{ij} are a set of $\tfrac{1}{2}N$ differences between pairs of integers chosen from the integers $1, 2, ..., N$. It is easily seen that Σr_{ij}^2 is largest when the pairs are $(1, N), (2, N-1), (3, N-2)$ and so on. In general

$$r_{ij} = (N-k+1) - k = N - 2k + 1 \quad (k = 1, 2, ..., \tfrac{1}{2}N). \tag{13}$$

so that

$$\Sigma r_{ij}^2 = \sum_{k=1}^{\tfrac{1}{2}N} (N - 2k + 1)^2 = \tfrac{1}{6}N(N^2 - 1), \tag{14}$$

and the supremum value of (12) is therefore

$$R^2(S_1) = \frac{N(N^2 - 1)}{6\pi} \sim \frac{N^3}{6\pi}. \tag{15}$$

We have denoted by S_1 the optimum S statistic

$$S_1 = \sum_{k=1}^{\tfrac{1}{2}N} (N - 2k + 1) h_{k, N-k+1},$$

for which

$$E(S_1 \mid \Delta = 0) = \tfrac{1}{2}\Sigma(N - 2k + 1) = \tfrac{1}{8}N^2,$$
$$D^2(S_1 \mid \Delta = 0) = \tfrac{1}{4}\Sigma(N - 2k + 1)^2 = \tfrac{1}{24}N(N^2 - 1).$$

(16)

The test based on S_1 is essentially a simplified version of Spearman's rank correlation test, which is in effect defined by

$$V = \sum_{i < j} (j - i) h_{ij},$$

(17)

where the summation extends over all possible $\tfrac{1}{2}N(N - 1)$ pairs of observations. Stuart (1954) has shown that

$$R^2(V) \sim \frac{N^3}{4\pi},$$

(18)

so that using (15) and (18) in (2) and (4) with $r = 3$, we obtain for the A.R.E. of S_1 compared to V

$$\text{A.R.E.}(S_1, V) = (\tfrac{3}{4})^{\frac{1}{2}} = 0.87.$$

(19)

The loss of A.R.E. involved in reducing the number of comparisons from $\tfrac{1}{2}N(N - 1)$ to $\tfrac{1}{2}N$ is as little as 13%.

These values of the A.R.E. depend on the assumption of normality, but the calculation of the form of the optimum statistic, S_1, and also of the statistic, S_3, of § 5, does not. For (7) remains true, with a changed numerical factor, for general continuous distributions.

5. An unweighted sign test

The relatively high efficiency of S_1 compared to V leads us to construct, by analogy, a simplified version of Kendall's rank correlation test, which may be defined by

$$Q = \sum_{i < j} h_{ij},$$

(20)

and gives equal weight to all $\tfrac{1}{2}N(N - 1)$ comparisons. Q has the same A.R.E. as V (Stuart, 1954). The analogous sign test, based on $\tfrac{1}{2}N$ equally weighted independent comparisons, is

$$S_2 = \Sigma h_{ij},$$

(21)

and using (10), we obtain, with all $w_{ij} = 1$,

$$R^2(S_2) = \frac{2}{N\pi} (\Sigma r_{ij})^2.$$

(22)

We now require to choose $\tfrac{1}{2}N$ pairs from the first N integers so that (22) or, equivalently, $\Sigma(j - i) = \Sigma r_{ij}$, takes its maximum value. This occurs whenever every i is chosen from the first $\tfrac{1}{2}N$ integers and every j from the last $\tfrac{1}{2}N$ integers. In particular, it occurs when every $r_{ij} = \tfrac{1}{2}N$ exactly, so that

$$S_2 = \sum_{k=1}^{\frac{1}{2}N} h_{k, \frac{1}{2}N+k},$$

(23)

and (22) becomes

$$R^2(S_2) = \frac{N^3}{8\pi}.$$

(24)

Using (24) and (18), we obtain

$$\text{A.R.E.}(S_2, V) = (\tfrac{1}{2})^{\frac{1}{2}} = 0.79,$$

(25)

while from (15)

$$\text{A.R.E.}(S_2, S_1) = (\tfrac{1}{2})^{\frac{1}{2}} = 0.91.$$

(26)

Thus the simplified version of Kendall's rank correlation test is 21 % less efficient than Q or V, and 9 % less efficient than the simplified Spearman coefficient S_1. The use of S_2 is equivalent to a test considered by Theil (1950).

6. THE BEST UNWEIGHTED SIGN TEST

However, we can improve on the efficiency of S_2, and in fact get very nearly as high an efficiency as that of S_1, by 'throwing away' some of the $\frac{1}{2}N$ comparisons and retaining equal weights for the others. This was suggested by one of the present authors in the discussion of Foster & Stuart (1954); it leads to an increase in efficiency because, by comparing observations further apart, individual comparisons are made more sensitive.

In (1), let every w_{ij} be either 0 or 1, and let m ($\leqslant \frac{1}{2}N$) be the number of non-zero w_{ij}. For our new statistic S_3 we have, as in (8),

$$E'(S_3) = -\frac{1}{2\sqrt{\pi}} \Sigma w_{ij} r_{ij} \quad (w_{ij} = 0 \text{ or } 1), \tag{27}$$

and from (9)
$$D^2(S_3 \mid \Delta = 0) = \tfrac{1}{4}m, \tag{28}$$

so that (3), (27) and (28) give

$$R^2(S_3) = \frac{1}{m\pi} (\Sigma w_{ij} r_{ij})^2 \quad (w_{ij} = 0 \text{ or } 1). \tag{29}$$

To maximize this by choice of m and r_{ij}, we again work in two stages. For fixed m, (29) will take its largest value when the comparisons given zero weights are based on the middle $(N - 2m)$ observations, while every i is chosen from the first m observations and every j is chosen from the last m observations. In particular, this will be so when every $r_{ij} = (N - m)$ exactly, so that

$$S_3 = \sum_{k=1}^{m} h_{k, N-m+k} \tag{30}$$

and (29) becomes
$$R^2(S_3) = \frac{m(N - m)^2}{\pi}. \tag{31}$$

(31) is the largest possible value of $R^2(S_3)$ for fixed m. (S_2 is the special case of S_3 when $m = \frac{1}{2}N$.) We now choose m to maximize (31). Differentiating, we get

$$m = \tfrac{1}{3}N \tag{32}$$

for a maximum, so that finally
$$S_3 = \sum_{k=1}^{\frac{1}{3}N} h_{k, \frac{2}{3}N+k},$$

for which
$$\left. \begin{aligned} E(S_3) &= \tfrac{1}{9}N, \\ V(S_3) &= \tfrac{1}{12}N, \end{aligned} \right\} \tag{33}$$

and from (31),
$$R^2(S_3) = \frac{4N^3}{27\pi}. \tag{34}$$

From (34) and (18), we have
$$\text{A.R.E. } (S_3, V) = (\tfrac{16}{27})^{\frac{1}{2}} = 0.84, \tag{35}$$

while from (15)
$$\text{A.R.E. } (S_3, S_1) = (\tfrac{8}{9})^{\frac{1}{2}} = 0.96. \tag{36}$$

Compared with either V or S_1, S_3 has about 5 % higher efficiency than S_2, and in fact its efficiency is 96 % of that of S_1, so that for practical purposes it may be recommended instead of S_1 because it requires no weighting of the comparisons.

7. Comparison of the sign tests

In Table 1, the A.R.E. of the sign tests are tabulated, compared to each other, to the rank correlation tests, and to the best (parametric) test against normal regression, based on the sample regression coefficient b, which has a value of (3) given by

$$R^2(b) \sim \frac{N^3}{12},\tag{37}$$

as follows immediately from the fact that b is an unbiased estimator of Δ with variance $12/\{N(N^2-1)\}$.

Table 1. *Asymptotic relative efficiencies of sign tests*

Test statistic	Asymptotic relative efficiency		
	Compared to S_1	Compared to rank correlation tests	Compared to best parametric test
$S_1 = \sum\limits_{k=1}^{\frac{1}{2}N} (N-2k+1)\,h_{k,\,N-k+1}$	1·00	0·87	0·86
$S_2 = \sum\limits_{k=1}^{\frac{1}{2}N} h_{k,\,\frac{1}{2}N+k}$	0·91	0·79	0·78
$S_3 = \sum\limits_{k=1}^{\frac{1}{2}N} h_{k,\,\frac{1}{2}N+k}$	0·96	0·84	0·83

From (2), (18) and (37), it follows that the A.R.E. of either rank correlation coefficient compared to b is

$$\text{A.R.E.}\,(V,b) = \left(\frac{3}{\pi}\right)^{\frac{1}{2}} = 0.98,\tag{38}$$

and not $3/\pi = 0.95$ as given by Stuart (1954).

8. Comparison with A.R.E. of other tests

Apart from the two rank correlation tests already discussed, Stuart (1954) investigated the A.R.E. of three other distribution-free tests for trend in location. Two of these, the rank serial correlation test and the turning point test, were found to have zero values of R as defined by (3); the third, the difference-sign test, was found to have a value of r equal to 1 in (4), as against $r = 3$ for all the tests considered in this paper. It followed that the three tests mentioned all have A.R.E. zero compared to the rank correlation tests (and hence to all the tests discussed here). Noether (1955) gives general results which rigorize these conclusions.

A well-known and simple test which has not, as far as we know, previously been discussed from the point of view of A.R.E. is the median test, due to Brown & Mood (1951). The N (even) observations are divided into two sets of $\frac{1}{2}N$ consecutive observations. The test

statistic is simply the number of observations in the first set which exceed the sample median y_m, and it is therefore defined by

$$B = \sum_{i=1}^{\frac{1}{2}N} b_{im}, \tag{39}$$

where

$$b_{im} = \begin{cases} 1 & \text{if } y_i > y_m, \\ 0 & \text{if } y_i < y_m. \end{cases}$$

The A.R.E. of B is easily obtained. We know that y_i is a normal variate with mean $(\alpha + i\Delta)$ and unit variance. It follows that the sample median y_m is asymptotically a normal variate with mean $(\alpha + \frac{1}{2}(N+1)\Delta)$ and variance of order N^{-1}. Since y_i and y_m are asymptotically independent, $(y_i - y_m)$ is asymptotically normal with mean $\Delta[i - \frac{1}{2}(N+1)]$ and unit variance, so that for $i < \frac{1}{2}(N+1)$

$$E(b_{im}) = \text{prob}\,(y_i > y_m) \sim 1 - G\{\Delta[\tfrac{1}{2}(N+1) - i]\}. \tag{40}$$

Using (6) in (40), we obtain

$$E'(b_{im}) \sim -\frac{1}{\sqrt{(2\pi)}}[\tfrac{1}{2}(N+1) - i], \tag{41}$$

so that from (39) and (41),

$$E'(B) = \sum_{i=1}^{\frac{1}{2}N} E'(b_{im}) \sim -\frac{1}{\sqrt{(2\pi)}} \sum_{i=1}^{\frac{1}{2}N} [\tfrac{1}{2}(N+1) - i] \sim -\frac{N^2}{8\sqrt{(2\pi)}}. \tag{42}$$

Also (Brown & Mood, 1951)

$$D^2(B \mid \Delta = 0) \sim \frac{N}{16}. \tag{43}$$

(42) and (43) give, in (3),

$$R^2(B) \sim \frac{N^3}{8\pi}. \tag{44}$$

Comparison of (44) with (24) shows that B has precisely the same A.R.E. as S_2, and is therefore slightly less efficient than S_1 and S_3. If the observations are available in serial order, S_3 is simpler to compute than B, which involves ranking all the observations to find the median, and then making $\frac{1}{2}N$ comparisons, as against $\frac{1}{3}N$ for S_3. There is therefore no reason to prefer B to S_3 in this case. If, however, the data were available graphically, B would be considerably easier to compute, and this would outweigh the slight loss of efficiency compared to S_3.

9. COMPARISON OF THE POWERS OF TESTS

So far we have compared tests by the A.R.E. in the usual way. Before considering the power of the test S_3 in small samples it is convenient to examine the meaning of the A.R.E. more carefully. If the A.R.E. of a quick test relative to an efficient test is A, then asymptotically A^{-1} as many observations have to be made for the quick test to give the same local power as the efficient test. This is directly relevant if in designing an experiment a choice has to be made between, on the one hand, using an efficient method of analysis and on the other taking more observations and using a quick method of analysis. But it is not directly relevant to the choice of a method of analysis for a given body of data, because it depends in part on r, defined by (4), measuring the rate at which power increases with increasing N. For a given problem r is fixed and so the A.R.E. can be reinterpreted in terms of the power attained at a fixed sample size, but it seems preferable to compare tests directly in terms of power.

Consider a test based on a statistic S normally distributed with mean $E(S \mid \Delta)$ and standard deviation $D(S \mid \Delta)$, where the null hypothesis is $\Delta = 0$. The null hypothesis is rejected at the significance level α if

$$S > E(S \mid 0) + \lambda_a D(S \mid 0), \tag{45}$$

where

$$G(-\lambda_a) = \alpha. \tag{46}$$

The power of the test is $G[p(\Delta)]$, where

$$p(\Delta) = \frac{E(S \mid \Delta) - E(S \mid 0) - \lambda_a D(S \mid 0)}{D(S \mid \Delta)}. \tag{47}$$

Now

$$p'(0) = \left(\frac{\partial p(\Delta)}{\partial \Delta}\right)_{\Delta = 0} = \frac{E'(S \mid 0) + \lambda_a D'(S \mid 0)}{D(S \mid 0)}. \tag{48}$$

In all the applications in this paper $D'(S \mid 0) = 0$, so that

$$p'(0) = \frac{E'(S \mid 0)}{D(S \mid 0)} = R(S). \tag{49}$$

Near $\Delta = 0$,

$$p(\Delta) = \Delta R(S) - \lambda_a + O(\Delta^2), \tag{50}$$

and in applications the first two terms give, asymptotically in N, the whole of the power curve. Moreover, $R(S) \sim RN^{\frac{1}{2}r}$ as $N \to \infty$ and comparable tests of a given hypothesis will have the same r; hence we usually need to consider just R. We call $p(\Delta)$ the *power deviate* and $p'(0)$ the *power derivative*. Asymptotically the graph of $p(\Delta)$ against Δ is linear, and tests at different significance levels are given by parallel straight lines. Or to put the same fact another way, the power curves are asymptotically linear when plotted on arithmetical probability paper.

Now consider the small sample theory with S possibly not normally distributed. Then if the power curves are plotted on probability paper they can be expected to form an approximately parallel set of curves approaching a set of parallel lines as the sample size increases and the distribution of S tends to normality. This is of course only a method of presenting the results of power calculations, but we shall find it very convenient both in assessing the small-sample behaviour and in comparing different tests.

Consider now two tests for which the asymptotic values of $R(S)$ are R_1, R_2. Then asymptotically in N the power curves for a given α are two lines on probability paper, the ratio of their slopes being R_1/R_2 independent of α; both lines intersect the probability axis at α.

A first consequence is that there is no simple general relation between the difference in the power of the two tests and the ratio R_1/R_2. If $R_1 \neq R_2$ we can, by taking α sufficiently small, make the difference in power between the two tests arbitrarily near unity. In practice we are probably only interested in $0 \cdot 20 \geqslant \alpha \geqslant 0 \cdot 001$, but the general conclusion remains that the difference in power between a quick test and an efficient test will be greatest for small α. Table 2 expresses this quantitatively; it shows for given R_1/R_2 the powers of the two tests at the point at which the difference in powers is greatest. The values in Table 2 are independent of N, but the values of Δ at which these powers are attained do depend on N. This is the restriction on the alternative hypothesis referred to in § 3. Thus if $R_1/R_2 = 0 \cdot 7$ and $\alpha = 0 \cdot 05$, the difference in power is greatest for the value of Δ at which the power of the efficient test is 77 % and of the quick test 51 %.

Now consider the power of S_3 in small samples. Two methods can be used. The first is to take the expansion (50) to higher powers of Δ and to introduce a correction for the non-normality of S based on an Edgeworth expansion. This may be shown to give good results even for very small N, and is a general method which could be used where direct numerical calculation is difficult. However, for S_3 it is much easier to calculate the power directly from the National Bureau of Standards tables of the binomial distribution (1950).

Table 2. *Asymptotic theory. Powers (per cent) of quick and efficient tests at points at which difference in power is greatest**

R_1/R_2 α	0·10	0·05	0·01	0·001
0·9	67, 73	63, 71	49, 60	54, 67
0·8	61, 74	56, 72	49, 71	43, 72
0·7	59, 80	51, 77	42, 77	39, 83
0·6	54, 84	47, 84	39, 86	29, 87
0·5	48, 88	41, 89	30, 90	20, 93
0·3	35, 96	27, 96	14, 97	7, 99

The power was computed in this way for $N = 15\,(15)\,135$, the significance level being the largest value $\leqslant 0{\cdot}05$. Under the null hypothesis the test statistic is distributed as $(\frac{1}{2} + \frac{1}{2})^{\frac{1}{3}N}$ and under the alternative hypothesis as $(p+q)^{\frac{1}{3}N}$, where

$$p = G\!\left(-\frac{\sqrt{2N\Delta}}{3}\right). \tag{51}$$

The power corresponding to given p, $\frac{1}{3}N$ can be read off directly from the tables and (51) solved for Δ. The results are given in Table 3. For comparative purposes the exact power of the parametric test based on the regression coefficient, b, has been computed for the same values of N and Δ. When the standard deviation about the regression line is known, the power is exactly $G[p(\Delta)]$, where

$$p(\Delta) = \{\tfrac{1}{12}N(N^2-1)\}^{\frac{1}{2}}\,\Delta - \lambda_\alpha. \tag{52}$$

To avoid rewriting the values of Δ in Table 4 the rows of both tables have been lettered, and each entry in Table 4 relates to the value of Δ shown above the corresponding entry in Table 3.

Asymptotically, the ratio of the R values of the two tests is, by (34) and (37), $4/(3\sqrt{\pi}) = 0{\cdot}75$; the interpretation of this in terms of power can be obtained from Table 2. The full curve in Fig. 1 and the full curves in Fig. 2 for $k = 0$ show the power curves for $N = 15, 30$, and the dotted lines are the corresponding asymptotic power curves. The small-sample power is lower than the value given by the asymptotic theory, the difference being quite appreciable in the region of 80–90 % power. The power curves of the most efficient test are exactly linear and differ from their asymptotic form only because of the very small difference between $\{N(N^2-1)\}^{\frac{1}{2}}$ and $N^{\frac{3}{2}}$. Hence the test S_3 is less efficient relative to b than the asymptotic

---

* These values were obtained graphically by drawing on probability paper lines whose ratio of slopes is R_1/R_2 and reading off the maximum difference in probability between them. The differences in power are determined accurately, but it is rather difficult to find the precise point of maximum difference. The values in Table 2 involve R_1, R_2 only through their ratio R_1/R_2.

Table 3. *Exact power of S_3 test against normal regression alternatives*

Values of the standardized regression coefficient, Δ, are given in parentheses, and the corresponding power appears immediately below.

Sample size (N) ...	15	30	45	60	75	90	105	120	135
Significance level α ...	0·031	0·011	0·018	0·021	0·022	0·049	0·045	0·040	0·036
a	(0·0035) 0·035	(0·0018) 0·013	(0·0012) 0·021	(0·0009) 0·026	(0·0007) 0·027	(0·0006) 0·062	(0·0005) 0·057	(0·0004) 0·053	(0·0004) 0·048
b	(0·0178) 0·050	(0·0089) 0·023	(0·0059) 0·042	(0·0044) 0·055	(0·0036) 0·064	(0·0030) 0·135	(0·0025) 0·134	(0·0022) 0·133	(0·0020) 0·130
c	(0·0358) 0·078	(0·0179) 0·046	(0·0119) 0·091	(0·0090) 0·126	(0·0072) 0·154	(0·0060) 0·291	(0·0051) 0·306	(0·0045) 0·317	(0·0040) 0·327
d	(0·0545) 0·116	(0·0272) 0·086	(0·0182) 0·173	(0·0136) 0·245	(0·0109) 0·306	(0·0091) 0·508	(0·0078) 0·542	(0·0068) 0·572	(0·0060) 0·598
e	(0·0742) 0·168	(0·0371) 0·149	(0·0247) 0·297	(0·0185) 0·416	(0·0148) 0·512	(0·0124) 0·730	(0·0106) 0·773	(0·0093) 0·807	(0·0082) 0·836
f	(0·0954) 0·237	(0·0477) 0·244	(0·0318) 0·461	(0·0238) 0·617	(0·0191) 0·727	(0·0159) 0·894	(0·0136) 0·924	(0·0119) 0·946	(0·0106) 0·961
g	(0·1190) 0·328	(0·0595) 0·376	(0·0397) 0·648	(0·0298) 0·804	(0·0238) 0·891	(0·0198) 0·974	(0·0170) 0·986	(0·0149) 0·992	(0·0132) 0·996
h	(0·1466) 0·444	(0·0733) 0·544	(0·0489) 0·823	(0·0366) 0·933	(0·0293) 0·975	(0·0244) 0·997	(0·0209) 0·999	(0·0183) 1·000	(0·0163) 1·000
i	(0·1812) 0·590	(0·0906) 0·736	(0·0604) 0·944	(0·0453) 0·989	(0·0362) 0·998	(0·0302) 1·000	(0·0259) 1·000	(0·0227) 1·000	(0·0201) 1·000
j	(0·2326) 0·774	(0·1163) 0·914	(0·0775) 0·995	(0·0582) 1·000	(0·0465) 1·000	(0·0388) 1·000	(0·0332) 1·000	(0·0291) 1·000	(0·0258) 1·000

Table 4. *Exact power of b test against normal regression alternatives*

Values of Δ are given in parentheses above the corresponding entry in Table 3.

Sample size (N) ...	15	30	45	60	75	90	105	120	135
Significance level α ...	0·031	0·011	0·018	0·021	0·022	0·049	0·045	0·040	0·036
a	0·036	0·013	0·023	0·027	0·030	0·066	0·062	0·057	0·054
b	0·059	0·030	0·056	0·074	0·088	0·179	0·182	0·183	0·183
c	0·103	0·074	0·143	0·201	0·249	0·429	0·457	0·481	0·502
d	0·171	0·157	0·300	0·416	0·509	0·722	0·764	0·799	0·827
e	0·267	0·294	0·519	0·673	0·775	0·919	0·945	0·962	0·973
f	0·395	0·485	0·746	0·877	0·941	0·988	0·994	0·997	0·999
g	0·551	0·699	0·911	0·975	0·993	0·999	1·000	1·000	1·000
h	0·772	0·880	0·984	0·998	1·000	1·000	1·000	1·000	1·000
i	0·879	0·977	0·999	1·000	1·000	1·000	1·000	1·000	1·000
j	0·979	0·999	1·000	1·000	1·000	1·000	1·000	1·000	1·000

theory suggests. The difference is greater for smaller α. The corresponding graphs to Figs. 1 and 2 for higher N show that for α in the range 0·01–0·05, the asymptotic theory applies well for $N \geqslant 60$.

The next thing is to investigate whether the form of S_3, involving the rejection of the middle third of the set of observations, can profitably be modified in small samples. Suppose that $(\frac{1}{3}N - 2k)$ observations are rejected so that the number of comparisons is $(\frac{1}{3}N + k)$; the exact power function can be worked out from the binomial tables as before, but an immediate comparison is not possible because the significance levels for different values of α cannot be made equal, except by the artificial device of randomized tests. However, if the curves are plotted on probability paper they are almost parallel for different α and an
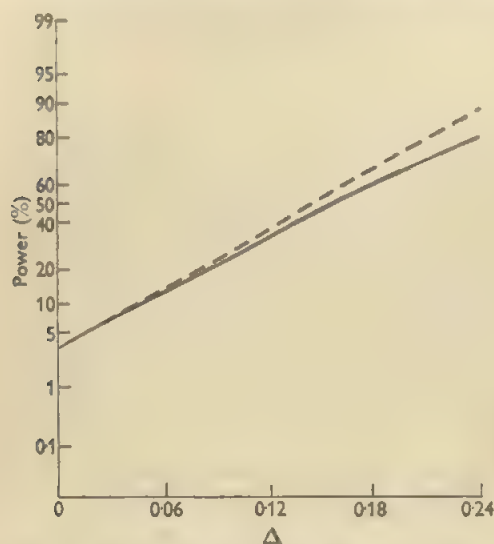
Fig. 1. Power of S_3 for $N = 15$. —— Exact power. –––– Value from asymptotic theory.
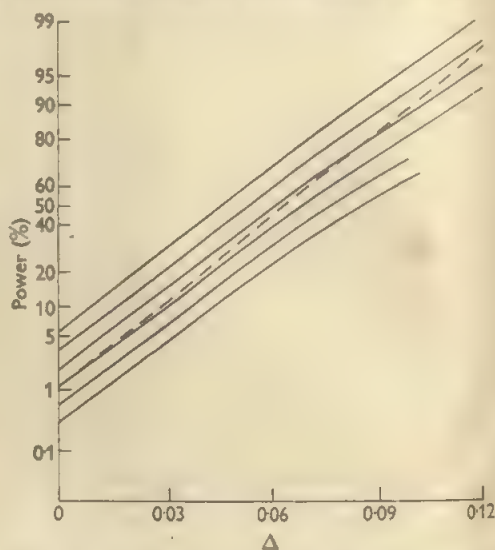
Fig. 2. Power of S_3 for $N = 30$. —— Exact power. –––– Value from asymptotic theory. Full curves are in descending order $k = 0, 1, 2, 0, 1, 2$, the second set having lower values of α than the first. Broken curve is $k = 0$.

increase in power with change in k would be shown by decreasing curvature. As would be expected, a negative k leads to a loss of power. Fig. 2 shows for $N = 30$ the curves for $k = 0, 1, 2$. There is a tendency for the curvature to increase as α decreases, but there does not seem to be any systematic change with k. Therefore, although an increase in k increases the number of available significance levels, it does not appreciably increase power. Hence there seems to be little value in modifying the $\frac{1}{3}$ rule in small samples; similar calculations for $N = 15$ confirm this.

We have not made the corresponding investigations for S_1.

10. EXAMPLES OF USE OF THE SIGN TESTS AGAINST TREND IN LOCATION

To illustrate the S_1 and S_3 tests, we use the figures of annual rainfall at Oxford for the years 1858–1952, quoted by Foster & Stuart (1954, Table 9).

For S_1 we compare the kth observation with the $(N - k + 1)$th, scoring 1 when the former is the larger and 0 when it is the smaller. The unit scores are then weighted by the distance

apart of the observations compared, i.e. by $(N - 2k + 1)$. In this case $N = 95$ and is odd, so that we must ignore the middle observation and proceed with $N = 94$. The unit scores are those with weights as follows:

89, 83, 79, 77, 75, 71, 65, 59, 57, 55, 51, 49, 47, 45, 33, 31, 27, 15, 3.

The value of S_1 is the sum of these weights, 1011. From (16), with $N = 94$, we have

$$E(S_1) = 1104.5, \quad D^2(S_1) = 34603.75, \quad D(S_1) = 186.0.$$

The observed value of S_1 thus represents a deviation from expectation of almost exactly one-half its standard error and is therefore in good agreement with the null hypothesis of zero trend.

If, alternatively, we were to use the simpler S_3 test we compare the kth observation with the $(\frac{2}{3}N + k)$th, scoring 1 or 0 as before, but no weighting is necessary. Since $N = 95$ and is not a multiple of three, we retain the extra observations (in accordance with our findings at the end of §9 above) and compare each of the first thirty-two observations with the corresponding observation in the last thirty-two. For our sequence of scores we obtain

1 0 1 0 1 0 0 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 1 0 0 0,

the total score S_3 being 14. This clearly agrees well with the expected value of $\frac{1}{2} \times 32 = 16$. The standard error of S_3 is, from (33), $\sqrt{(\frac{1}{4} \times 32)} = 2.83$, so that the deviation from expectation, corrected for continuity, is just over one-half of a standard error.

11. SIGN TESTS FOR TREND IN DISPERSION

We consider now tests for a trend not in location but in the dispersion about a fixed location. For example, in a regression problem we may want to test quickly whether the scatter about the regression curve increases as the independent variable increases.

Divide the series $x_1, \ldots, x_N$ into sets $x_1, \ldots, x_k; x_{k+1}, \ldots, x_{2k}; \ldots,$ rejecting a few observations in the centre of the original series if N is not exactly divisible by k. The best choice of k is discussed below. For each set of k observations find the range, w, thus getting a series of ranges $w_1, \ldots, w_r$, where r is the integral part of N/k. The ranges are then tested for trend by one or other of the tests S_1 and S_3.

If the null hypothesis is that $x_1, \ldots, x_N$ are independently distributed with constant dispersion about a regression line, $w_1, \ldots, w_r$ are independent and identically distributed. If the regression is not linear the w's will be approximately identically distributed unless the trend within sets of observations varies appreciably.

In the next section, a valid test is obtained for any k, and the best value of k for detecting certain special forms of trend is found for large samples. The behaviour for small samples has not been investigated. The following provisional rules are suggested:

$$\begin{aligned}
&\text{if} &N &\geqslant 90 &\text{take} \quad k = 5, \\
&\text{if} \quad 90 > N &\geqslant 64 &&\text{take} \quad k = 4, \\
&\text{if} \quad 64 > N &\geqslant 48 &&\text{take} \quad k = 3, \\
&\text{if} \quad 48 > N &&&\text{take} \quad k = 2.
\end{aligned}$$

Except when N is very large it is probably advisable to use the weighted, rather than the unweighted, sign test.

12. Choice of k for dispersion tests

To investigate the theory of the test for trend in dispersion we take a special form for the null and alternative hypotheses. Suppose that $x_1, ..., x_N$ are independently normally distributed with constant mean and with standard deviations $\sigma(1), ..., \sigma(N)$, where $\sigma(n)$ varies at most slowly with n. Then the ranges $w_1, ..., w_r$ defined in §11 have very nearly the distribution of ranges of k observations drawn from normal populations of standard deviations $\sigma_1 = \sigma(\frac{1}{2}k)$, $\sigma_2 = \sigma(\frac{3}{2}k)$, ... Patnaik (1950) has shown that a range of k observations can be represented to a close approximation as a multiple of a χ-variate with suitably chosen degrees of freedom, ν_k. Therefore $w_i^2 \sigma_j^2 / (w_j^2 \sigma_i^2)$ is approximately an F variate with (ν_k, ν_k) degrees of freedom.

Hence

$$\text{prob}\,(w_i/w_j > 1) \simeq \int_0^{(\sigma_i/\sigma_j)^2} \frac{\Gamma(\nu_k)}{[\Gamma(\frac{1}{2}\nu_k)]^2} \frac{x^{\frac{1}{2}\nu_k - 1}}{(1+x)^{\nu_k}}\, dx$$

$$\simeq \frac{1}{2} + \left\{ \left(\frac{\sigma_i}{\sigma_j} \right)^2 - 1 \right\} A_k$$

where

$$A_k = \frac{\Gamma(\nu_k)}{[\Gamma(\frac{1}{2}\nu_k)]^2} (\tfrac{1}{2})^{\nu_k},$$

provided that $(\sigma_i/\sigma_j)^2 - 1$ is small.

If we assume that the trend in standard deviation is such that $\sigma(n) = \sigma_0 e^{\gamma n} \sim \sigma_0 (1 + \gamma n)$, where $\gamma N \ll 1$, so that γ is the fractional increase in standard deviation per observation, we have

$$(\sigma_i/\sigma_j)^2 - 1 \sim 2k\gamma(i-j)$$

and

$$\text{prob}\,(w_i/w_j > 1) \simeq \tfrac{1}{2} + 2k\gamma(i-j)\,A_k.$$

Consider first the application to the ranges of the unweighted sign test, S_3. From the $r \simeq N/k$ ranges we make approximately $\frac{1}{3}N/k$ independent comparisons in each of which $i - j \simeq \frac{2}{3}r$. Therefore if S is the total score, its mean and standard deviation are given by

$$E(S \mid \gamma) \simeq \frac{1}{3}\frac{N}{k}\,4k\gamma\,\frac{2rA_k}{3} = \frac{8N^2 A_k \gamma}{9k},$$

$$D(S \mid \gamma) = \left(\frac{N}{3k} \right)^{\frac{1}{2}} + O(\gamma^2).$$

Therefore the power derivative, $p_3'(0)$, of the test is

$$p_3'(0) = \frac{E'(S \mid 0)}{D(S \mid 0)} \simeq \frac{8\sqrt{3}\,N^{\frac{3}{2}} A_k}{9\sqrt{k}}. \tag{53}$$

An exactly analogous calculation for the weighted sign test S_1 gives

$$p_1'(0) \simeq \frac{2\sqrt{6}\,N^{\frac{3}{2}} A_k}{3\sqrt{k}}. \tag{54}$$

Thus in both cases the asymptotically best value of k is the one that maximizes $A_k/\sqrt{k}$. From Patnaik's table of ν_k the values in Table 5 have been computed.

Now the number of ranges is N/k and the number of comparisons is one-half or one-third of this and is therefore small even when N is, by usual standards, quite large. Therefore it is advisable to use smaller values of k than the theoretical optimum in large samples. In the absence of an investigation of the small-sample properties of the test, the rule of §11

Table 5. *Determination of efficiencies of different set sizes, k, for testing trend in dispersion*

k	$A_k/\sqrt{k}$	k	$A_k/\sqrt{k}$
2	0·112	6	0·167
3	0·141	7	0·169
4	0·158	8	0·170
5	0·164	9	0·169

is suggested. This is based on the considerations that there is little gain in taking $k > 5$ and that it is advisable, whenever possible, to have at least sixteen ranges.

If we substitute, in (53) and (54), the value $A_k/\sqrt{k} \simeq 0·16$, we have

$$\left.\begin{aligned} p_3'(0) &\simeq 0·246 N^{\frac{3}{2}}, \\ p_1'(0) &\simeq 0·261 N^{\frac{3}{2}}. \end{aligned}\right\} \tag{55}$$

It remains to compare (55) with the corresponding quantity for the maximum-likelihood test of the corresponding parametric hypotheses.

13. A.R.E. OF DISPERSION TESTS

For simplicity assume that $x_1, \ldots, x_N$ are normally and independently distributed with zero mean and that the standard deviation of x_n is $\sigma_0 e^{\gamma n}$, where γN is small. The log likelihood is

$$L = -\tfrac{1}{2} N \log(2\pi) - N \log \sigma_0 - \gamma \sum_1^N n - \frac{1}{2\sigma_n^2} \Sigma x_n^2 e^{-2\gamma n}.$$

If we differentiate and take expectations, retaining only the terms independent of γ, and letting N tend to infinity, we get

$$E\left(\frac{\partial^2 L}{\partial \sigma_0^2}\right) \sim -\frac{2N}{\sigma_0^2}, \quad E\left(\frac{\partial^2 L}{\partial \sigma_0 \partial \gamma}\right) \sim -\frac{N^2}{\sigma_0}, \quad E\left(\frac{\partial^2 L}{\partial \gamma^2}\right) \sim -\tfrac{2}{3} N^3. \tag{56}$$

The large-sample variance of $\hat{\gamma}$, the maximum-likelihood estimate of γ, is given by inverting the Hessian matrix with elements (56). We get when γ is small and N is large

$$V(\hat{\gamma}) \sim 6/N^3. \tag{57}$$

Thus the power derivative of the test based on the maximum-likelihood estimate is

$$p_m'(0) = \frac{N^{\frac{3}{2}}}{\sqrt{6}} = 0·408 N^{\frac{3}{2}}. \tag{58}$$

(58) still applies if the mean in unknown or if a linear trend in mean has to be estimated.

From the formulae (55) and (58), and the fact that $p_x'(0) = R(x)$, it follows, on using (2) with $r = 3$, that the A.R.E.'s of the tests S_3, S_1 compared with the maximum-likelihood test are about 71 and 74 % respectively.

A test entirely analogous to the above tests can be found by calculating the variances within each set of k instead of the range. This is slightly more efficient in the parametric case but much of the simplicity of the test is lost and the increase in power may be shown to be trivial.

14. Examples of the use of sign tests against trend in dispersion

We again use for illustrative purposes the rainfall data quoted by Foster & Stuart (1954, Table 9). Using the provisional rule given in § 11 above, we take ranges of sets of five observations. Since $N = 95$, this gives us exactly nineteen sets, no rejection of observations being necessary. The nineteen ranges are:

9·64, 12·30, 12·01, 11·45, 5·43, 13·05, 9·86, 10·89, 6·95, 15·03,

11·34, 6·63, 12·19, 8·55, 4·80, 11·00, 7·76, 7·03, 10·98.

If we apply the test S_1 we drop the middle value and take the signs of 10·98–9·64, 7·03–12·30, ... down to 11·34–6·95, thus obtaining

score:	0	1	1	1	1	1	0	1	0
weight:	17	15	13	11	9	7	5	3	1

The total score is therefore 58, and from (16) with $N = 18$ the mean score is 40·5 and the variance is 242·25, so that the standard error is 15·6. The deviation from expectation is about 1·12 standard errors, and so the two-sided normal significance level is about 27 %. The exact significance level is, by enumeration, $73/256 \simeq 28\frac{1}{2}$ %.

If we use the test S_3 we reject the middle five of the nineteen ranges and take the signs of 12·19–9·64, etc. This gives scores

0　1　1　1　0　1　0

There is clearly good agreement with an equal probability for zeros and ones; significance would be tested in the binomial distribution $(\frac{1}{2} + \frac{1}{2})^7$. The test S_3 is not to be recommended in the present instance because with only seven comparisons the loss of sensitivity compared to the S_1 test would be considerable.

Thus although there is a slight indication that the dispersion decreases with time, both tests suggest that this could easily be a sampling fluctuation.

15. Sequential tests

Finally, we point out the possibility of constructing sequential tests for trend related to the tests considered above. While this paper was in preparation an abstract (Noether, 1954) appeared describing briefly a test rather similar to the one we had developed. Hence a full discussion will not be attempted here. However, some calculations in a special case suggest that the average sample size under the null and alternative hypotheses are, for the sequential sign tests, only a little greater than the corresponding parametric fixed sample size.

Sequential sign tests for trend are only likely to be of practical value under rather exceptional circumstances. For they require that observations are sufficiently easy to obtain for it to be worth while to use inefficient methods of analysis, and yet sufficiently difficult to obtain for the saving from the use of a sequential method to be important. A possible application is to the marking of a large number of examination scripts. If they are marked in alphabetical order it may be useful to test, as the marking proceeds, for a trend in the marks, which would indicate a changing standard of marking. A sequential method is appropriate and yet elaborate calculations would be out of place.

16. General comments

The calculation of the efficiency of the above tests and the determination of optimum weightings. etc.. has been based on a particular type of alternative hypothesis. It is clear in a general way that the tests will remain effective for detecting monotone trends. Positive serial correlation among the observations would increase the chance of a significant answer even in the absence of a trend.

The occurrence of ties has been ignored in the above work. A small number of ties can be dealt with by counting one-half a comparison in each direction. i.e. if $y_i = y_j$ we calculate as if one-half a comparison has $y_i > y_j$ and one-half has $y_i < y_j$. If a substantial proportion of the comparisons are ties a special investigation is necessary or the comparisons should be randomized.

Estimates for the trend could be constructed from the test statistics S_1. S_3. It is very doubtful if such estimates would be of value; in any case, in much work with quick tests, if the trend is shown to be significant it can be estimated graphically.

REFERENCES

Brown, G. W. & Mood, A. M. (1951). On median tests for linear hypotheses. *Proceedings of the Second Berkeley Symposium*, pp. 159–66. University of California Press.

Foster, F. G. & Stuart, A. (1954). Distribution-free tests in time-series based on the breaking of records. *J. R. Statist. Soc.* Series B (Methodological), **16**, 1–22.

National Bureau of Standards (1950). *Tables of the Binomial Probability Distribution*, Applied Mathematics Series, no. 6. Washington.

Noether, Gottfried, E. (1954). Abstract of 'A sequential test of randomness against linear trend'. *Ann. Math. Statist.* **25**, 176.

Noether, Gottfried E. (1955). 'The asymptotic relative efficiencies of tests of hypotheses'. *Ann. Math. Statist.* (to be published).

Patnaik, P. B. (1950). The use of mean range as an estimator of variance in statistical tests. *Biometrika*, **37**, 78–87.

Pitman, E. J. G. (1948). Lecture notes on 'Non-parametric statistical inference', University of North Carolina Institute of Statistics (mimeographed).

Stuart, Alan (1954). The asymptotic relative efficiencies of distribution-free tests of randomness against normal alternatives. *J. Amer. Statist. Ass.* **49**, 147–57.

Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. *Indag. math.* **12**, 85–91, 173–7.

THE VARIANCE OF THE MAXIMUM OF PARTIAL SUMS OF A FINITE NUMBER OF INDEPENDENT NORMAL VARIATES

By A. A. ANIS

Cairo University

1. THE PROBLEM

Consider n independent standard normal variates $X_1, X_2, ..., X_n$ and their partial sums

$$S_r = X_1 + ... + X_r \quad (r = 1, 2, ..., n).$$

Let
$$U_n = \underset{r}{\text{Max}} \{S_r\}$$

denote the maximum of these partial sums.

In a paper by Anis & Lloyd (1953) the expectation of U_n was studied and it was shown that

$$\mathscr{E}(U_n) = (2\pi)^{-\frac{1}{2}} \sum_1^{n-1} r^{-\frac{1}{2}}.$$

In the present paper the second moment about zero and hence the variance of U_n is obtained (see equation (7·1)).

We shall always use the symbol $\phi(x)$ to denote the probability density function of the standard normal variate, i.e.
$$\phi(x) = (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2),$$

and $F_n(x), f_n(x)$ to denote respectively the distribution function and the probability density function of U_n:
$$F_n(x) = Pr(U_n \leqslant x), \quad f_n(x) = \frac{d}{dx} F_n(x).$$

We have
$$F_n(y) = \int (n) \int_K \prod_1^n \phi(x_i)\, dx_i \quad (-\infty \leqslant y \leqslant \infty), \tag{1·1}$$

where the region of integration K is defined by

$$K: \quad \sum_1^r x_i \leqslant y \quad (r = 1, 2, ..., n).$$

It may be deduced that
$$F_n(x) = \int_0^\infty F_{n-1}(t)\, \phi(x - t)\, dt, \tag{1·2}$$

and that
$$f_n(x) = F_{n-1}(0)\, \phi(x) + \int_0^\infty f_{n-1}(t)\, \phi(x - t)\, dt. \tag{1·3}$$

2. THREE LEMMAS ON THE $F_r(0)$

At this stage we state three results relating to the $F_r(0)$ which we shall need in the sequel.

LEMMA 1.
$$\sum_{s=0}^n F_s(0)\, F_{n-s}(0) = 1. \tag{2·1}$$

This was proved in Anis & Lloyd (1953), and is repeated here merely for completeness.

LEMMA 2. $\qquad\qquad F_r(0) = (2r)!/2^{2r}(r!)^2.$ $\qquad\qquad$ (2·2)

To prove this we define a generating function

$$M(\lambda) = \sum_{i=0}^{n} \lambda^i F_i(0).$$

Then, using Lemma 1, it is readily seen that

$$M^2(\lambda) = (1-\lambda)^{-1}.$$ $\qquad\qquad$ (2·3)

Picking out the appropriate coefficient from $M(\lambda) = (1-\lambda)^{-\frac{1}{2}}$ gives the required result.

LEMMA 3. $\qquad\qquad \sum_{r=0}^{n} r F_r(0) F_{n-r}(0) = \tfrac{1}{4}n.$

This follows from Lemma 2 on differentiating (2·3) and equating coefficients of λ^{n-1}.

3. THE SECOND MOMENT OF U_n AS A LINEAR COMPOUND OF THE $F_r(0)$

The second moment $\mu_2(n)$ of U_n is given by

$$\mu_2(n) = \int_{-\infty}^{\infty} x^2 f_n(x)\, dx,$$

and, using the reduction formula (1·3), this becomes

$$\mu_2(n) = F_{n-1}(0) + \int_{-\infty}^{\infty}\int_{0}^{\infty} x^2 \phi(x-t) f_{n-1}(t)\, dt.$$

The double integral can be integrated once, with respect to x. Using well-known properties of $\phi(x)$, and remembering that F_{n-1} is the integral function of f_{n-1}, we obtain

$$\mu_2(n) = 1 + \int_{0}^{\infty} t^2 f_{n-1}(t)\, dt.$$ $\qquad\qquad$ (3·1)

We now use the reduction formula (1·3) a second time, resulting in

$$\mu_2(n) = 1 + F_{n-2}(0) \int_{0}^{\infty} t^2 \phi(t)\, dt + \int_{0}^{\infty}\int_{0}^{\infty} t^2 \phi(x-t) f_{n-2}(t)\, dt.$$

The last integral may be reduced in the same way. Continuing, we find

$$\mu_2(n) = 1 + \sum_{r=1}^{n-1} g_r F_{n-r-1}(0),$$ $\qquad\qquad$ (3·2)

where $\qquad g_r = \int_{0}^{\infty}(r)\int_{0}^{\infty} y_1^2 \phi(y_1-y_2)\,\phi(y_2-y_3)\cdots\phi(y_{r-1}-y_r)\,\phi(y_r)\prod_{1}^{r} dy_i.$ $\qquad$ (3·3)

4. THE COEFFICIENTS g_r OF THE LINEAR COMPOUND

We now seek to evaluate the g_r. As a first step we make the transformation

$$\left.\begin{aligned} x_i &= y_i - y_{i+1} \quad (i = 1, 2, \ldots, r;\ y_{r+1} = 0), \\ y_s &= \sum_{s}^{r} x_i \quad\quad (s = 1, 2, \ldots, r). \end{aligned}\right\}$$ $\qquad$ (4·1)

or

We then have
$$y_r = \int(r)\int(x_1+x_2+\ldots+x_r)^2 \prod_1^r \phi(x_i)\,dx_i,$$
(4·2)

where the region of integration is
$$R: \quad \sum_s x_i \geqslant 0 \quad (s=1,2,\ldots,r).$$

Hence
$$y_r = L_r + H_r,$$

where
$$L_r = \int(r)\int\left(\sum_1^r x_i^2\right)\prod_1^r \phi(x_i)\,dx_i, \quad H_r = \int(r)\int\left(\sum_{i \neq j}\sum x_i x_j\right)\prod_1^r \phi(x_i)\,dx_i.$$
(4·3)

We now consider these two integrals separately.

5. Evaluation of L_r

The integral L_r of (4·3) is readily evaluated, as follows. Since the integrand is spherically symmetrical the value of the integral is proportional to the magnitude of the r-dimensional solid angle defined by the region of integration R; and this in turn is proportional to the $F_r(0)$, since by (1·1) we have
$$F_r(0) = \int(r)\int \prod_1^r \phi(x_i)\,dx_i.$$
(5·1)

Let us write $F_r^0(0)$ to denote the integral of the integrand of $F_r(0)$ taken over the whole r-space, and let L_r^0 be similarly related to L_r; then the proportionality of our functions $F_r(0)$ and L_r to the solid angle R enable us to write
$$F_r(0):F_r^0(0) = L_r:L_r^0$$
whence
$$L_r = L_r^0 F_r(0),$$
(5·2)
since
$$F_r^0(0) = 1.$$

The value of L_r^0 is easily obtained by considering the standard integral
$$(2\pi)^{-\frac12 r}\int_{-\infty}^{\infty}(r)\int_{-\infty}^{\infty}\exp\left(-\tfrac12 k\sum_1^r x_i^2\right)\prod_1^r dx_i = k^{-\frac12 r}.$$

If we differentiate this expression with respect to k and then put $k=1$ we find
$$L_r^0 = r.$$

Hence, from (5·2),
$$L_r = rF_r(0).$$
(5·3)

The $F_r(0)$ are, of course, known explicitly.

6. Evaluation of H_r

We now consider the integral H_r of (4·3). We first transform back to the original variables y_i by the transformation (4·1). In this process the cross-product term transforms as follows:
$$\tfrac12\sum_{i \neq j}\sum x_i x_j = \sum_1^{r-1}\left\{x_i \sum_i^r x_j\right\} = \sum_1^{r-1}(y_i - y_{i+1})y_{i+1}$$
$$= \sum_{s=1}^{r-1}\left\{(2y_s - y_{s-1} - y_{s+1})\sum_{s+1}^r y_i\right\},$$

provided we interpret y_0 and y_{r+1} conventionally as

$$y_0 = y_1, \quad y_{r+1} = 0. \tag{6.1}$$

Expressing H_r in terms of the y's, and retaining these conventions, we then have

$$\tfrac{1}{2}H_r = \sum_{s=1}^{r-1} K_s(r), \tag{6.2}$$

where

$$K_s(r) = \int_0^\infty (r) \int_0^r (2y_s - y_{s-1} - y_{s+1}) \sum_{s+1}^r y_i \prod_1^r \phi(y_i - y_{i-1}) \, dy_i. \tag{6.3}$$

The reason for writing $K_s(r)$ in this form is that it enables us to perform one of the integrations at once, since

$$\int_0^\infty (2y_s - y_{s-1} - y_{s+1}) \, \phi(y_{s-1} - y_s) \, \phi(y_s - y_{s+1}) \, dy_s = \phi(y_{s-1}) \, \phi(y_{s+1}). \tag{6.4}$$

as is immediately seen on using the explicit form of the functions ϕ.

Equation (6.2) then becomes

$$K_s(r) = \int_0^\infty (s-1) \int_0^\infty \prod_1^{s-2} \phi(y_i - y_{i-1}) \, \phi(y_{s-1}) \, dy_1 \dots dy_{s-1}$$

$$\times \int_0^\infty (r-s) \int_0^\infty \left(\sum_{s+1}^r y_i \right) \phi(y_{s+1}) \prod_{s+1}^r \phi(y_i - y_{i-1}) \, dy_{s+1} \dots dy_r. \tag{6.5}$$

Now in this expression the $(s-1)$-fold integral is equal to $F_{s-1}(0)$, as may be seen on applying the transformation (4.1) to (5.1).

The other factor in the last expression, an $(r-s)$-fold integral, we call G_{r-s}. Thus

$$G_k = \int_0^\infty (k) \int_0^\infty \left(\sum_1^k y_i \right) \phi(y_1) \, \phi(y_1 - y_2) \dots \phi(y_{k-1} - y_k) \, \phi(y_k) \, dy_1 \dots dy_k. \tag{6.6}$$

We now proceed to show that this is expressible in terms of an integral previously evaluated by Anis & Lloyd (1953):

$$E_s = \int_0^\infty (s) \int_0^\infty \phi(y_1) \, \phi(y_1 - y_2) \dots \phi(y_{s-1} - y_s) \, \phi(y_s) \, dy_1 \dots dy_s = (2\pi)^{-\frac{1}{2}} (s+1)^{-\frac{1}{2}}. \tag{6.7}$$

We use the identity

$$\sum_1^k y_r = \sum_1^k \tfrac{1}{2} r(k+1-r) \, (2y_r - y_{r-1} - y_{r+1}), \tag{6.8}$$

where, by convention, we take

$$y_0 = y_{k+1} = 0.$$

Equation (6.6) becomes

$$G_k = \sum_{r=1}^k \tfrac{1}{2} r(k-r+1) \int_0^\infty (k) \int_0^\infty (2y_r - y_{r-1} - y_{r+1}) \, \phi(y_1) \prod_1^{k-1} \phi(y_i - y_{i+1}) \, \phi(y_k) \, dy_1 \dots dy_k. \tag{6.9}$$

We now use (6.3) to carry out one of the integrations, thus reducing the k-fold integral to

$$\int_0^\infty (k-1) \int_0^\infty \phi(y_1) \, \phi(y_1 - y_2) \dots \phi(y_{r-2} - y_{r-1}) \, \phi(y_{r-1}) \, \phi(y_{r+1}) \, \phi(y_{r+1} - y_{r+2})$$

$$\dots \phi(y_{k-1} - y_k) \, \phi(y_k) \, dy_1 \dots dy_{r-1} dy_{r+1} \dots dy_k$$

$$= E_{r-1} E_{k-r}. \tag{6.10}$$

Gathering the results of this section together, we have from (6.2), (6.5), (6.9) and (6.10),

$$H_r = 2 \sum_{s=1}^{r-1} K_s(r), \quad K_s(r) = F_{s-1}(0) G_{r-s}, \quad G_k = \sum_{r=1}^k \tfrac{1}{2} r(k+1-r) E_{r-1} E_{k-r}, \quad E_r = (2\pi)^{-\frac{1}{2}} (r+1)^{-\frac{1}{2}}. \tag{6.11}$$

7. CONCLUSION

We have found (3·2) that the second moment of the maximum of the partial sums $X_1, X_1 + X_2, ..., X_1 + ... + X_n$ is

$$\mu_2'(n) = 1 + \sum_{r=1}^{n-1} g_r F_{n-r-1}(0),$$

where

$$g_r = rF_r(0) + 2\sum_{s=1}^{n-1} F_{s-1}(0)\, G_{r-s}.$$

Thus

$$\mu_2'(n) = 1 + \sum_{r=1}^{n-1} rF_r(0)\, F_{n-r-1}(0) + 2\sum_{r=1}^{n-1}\sum_{s=1}^{r-1} F_{n-r-1}(0)\, F_{s-1}(0)\, G_{r-s}.$$

Short table of first and second moments about zero, and standard deviation, of the maximum of the partial sums of n independent standard normal deviates

n	μ_1'	μ_2'	σ
3	0·6810	2·1592	1·3021
4	0·9114	2·8842	1·4311
5	1·1108	3·6476	1·5536
6	1·2893	4·4367	1·6657
7	1·4521	5·2446	1·7709
8	1·6029	6·0671	1·8702
9	1·7440	6·9013	1·9647
10	1·8769	7·7451	2·0548
11	2·0031	8·5971	2·1412
12	2·1234	9·4560	2·2242
13	2·2385	10·3210	2·3043
14	2·3492	11·1914	2·3817
15	2·4558	12·0665	2·4567
16	2·5588	12·9459	2·5295
17	2·6585	13·8292	2·6002
18	2·7553	14·7160	2·6691
19	2·8493	15·6060	2·7363
20	2·9409	16·4989	2·8018
21	3·0301	17·3946	2·8659
22	3·1171	18·2928	2·9285
23	3·2022	19·1934	2·9899
24	3·2854	20·0962	3·0500
25	3·3668	21·0010	3·1090

The second term on the right-hand side equals $\frac{1}{2}(n-1)$, by Lemma 3, and the third term reduces, with the aid of Lemma 1, to $\sum_1^{n-2} G_r$. So we finally obtain

$$\mu_2'(n) = \frac{1}{2}(n+1) + 2\sum_1^{n-2} G_r,$$

where, by (6·11),

$$G_r = \frac{1}{2} \cdot \frac{1}{2\pi} \sum_{j=1}^{r} \{j(r-j+1)\}^{-\frac{1}{2}}.$$

Finally, then $\qquad \mu_2'(n) = \frac{1}{3}(n+1) + \frac{1}{2\pi} \sum_{r=1}^{n-2} \sum_{s=1}^{r} \{s(r-s+1)\}^{-\frac{1}{2}}.$ $\qquad$ (7·1)

From the point of view of numerical evaluation it is fortunate that the individual terms of the summand in (7·1) are independent of n; computations carried out for a given value of n can be immediately utilized for larger values of n. A short table of specimen values is appended.

Values corresponding to very high n may be approximated by a result of Erdös & Kac (1946), who gave the limiting distribution of $\theta_n = n^{-\frac{1}{2}} U_n$ as

$$\lim_{n \to \infty} \mathrm{Pr}\{\theta_n \leqslant \alpha\} = \sqrt{\frac{2}{\pi}} \int_0^\alpha \exp\left(-\tfrac{1}{2}x^2\right) dx \quad (\alpha \geqslant 0)$$

$$= 0 \qquad\qquad\qquad (\alpha < 0).$$

The limiting second moment of θ_n is thus unity, and the asymptotic second moment of U_n is n.

Our results are in agreement with this. We may approximate the double sum in (7·1) by the double integral

$$\int_{y=\frac{1}{2}}^{n-\frac{3}{2}} \int_{x=\frac{1}{2}}^{y} x^{-\frac{1}{2}}(y+1-x)^{-\frac{1}{2}} dx\, dy.$$

If we reverse the order of integration, this may easily be evaluated explicitly to give an expression which, to terms of order $n^{\frac{1}{2}}$, reduces to

$$n\pi - 2(2 + \sqrt{2})\, n^{\frac{1}{2}}.$$

Thus $\qquad\qquad\qquad\qquad \mu_2' \sim n - \frac{2 + \sqrt{2}}{\pi} \sqrt{n}.$

For $n = 25$ this gives $\mu_2' \sim 19\cdot6$; the correct value is $\mu_2' = 21\cdot001$.

The author wishes to acknowledge his debt to the referee for the asymptotic results given in the last paragraph.

REFERENCES

Anis, A. A. & Lloyd, E. H. (1953). On the range of partial sums of a finite number of independent normal variates. *Biometrika*, **40**, 35.

Erdös, P. & Kac, M. (1946). On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.* **52**, 292.

SPATIAL POINT PROCESSES, WITH APPLICATIONS TO ECOLOGY

By H. R. THOMPSON

Applied Mathematics Laboratory, D.S.I.R., Wellington, New Zealand

1. INTRODUCTION

The term 'point processes', referring to stochastic processes in which events occur at more or less irregular intervals and which are represented by points on the time-axis, is of comparatively recent origin, although the existence of such processes has in fact been well known for a long time. They have been discussed fairly extensively in such diverse applications as the counting of radioactive impulses, telephone calls and cases of contagious diseases. Wold (1949) developed a statistical theory for treating processes of this type, and also mentioned briefly how the events could take place in a two-dimensional or higher field. Such a generalization, from events with no time extension to those with no 'space' extension (i.e. specifically of a point character), has a suitable field of application in the ecological study of the distributional pattern of plants. If we can assume to a first approximation that the plants have the dimensions of a point, then we shall see that it is possible to discuss precisely probability relationships between the numbers of plants in different areas of the region under investigation.

The main aims of quantitative ecology are the precise description of a community of plants with interpretations in terms of the biology of the species, and the correlation of vegetational and environmental data, and ecologists have used several methods in an attempt to achieve these aims. In most of the initial work on field sampling for ecological data, the procedure was to take 'quadrats' (sample areas small in relation to the total area of the region) scattered at random over the area, and study statistics derived from the frequency distribution of the numbers of plants per quadrat. While this approach is useful to some extent, in that any given type of distribution function may be fitted to the data, it does not necessarily furnish the kind of information required by an ecologist. It will not give any evidence of trends, or indicate the pattern of the distribution over the area or the way in which this pattern may have arisen, all factors of prime importance in the study of the structure of a plant community. We only have to cite the negative binomial distribution, which is known to arise in at least four different ways, all based on widely differing assumptions, to illustrate this point.

In recent years ecologists have become aware of the need for a more satisfactory approach to the problem, and Greig-Smith (1952) provided a potentially great advance on the statistical side when he recommended the use of a grid of contiguous quadrats over some portion or portions of the region. The advantage, of course, in arranging the quadrats in a grid is that the analysis of variance technique may be employed, either for the detection of trends, or, more importantly, for the detection of a mosaic variation in density (due to ecological causes connected with the spread of the plants) by a 'nested sampling' type of analysis of variance, associating the quadrats into successively larger blocks and comparing the component block variances. The details and applications of this method are described at length by Greig-Smith, together with the results from sampling experiments on artificial

plant communities. We are not concerned here with discussing the ecological implications of the method (for which the reader is referred to a paper to appear elsewhere (Thompson, 1955)), but with the application of point process techniques to deriving the probability relations between the numbers of plants in the quadrats of a grid, required particularly in the analysis of variance and its sampling theory.

Wold's treatment of one-dimensional point processes is based on the distribution of the time interval between successive events, which could be either independent, or dependent on the previous interval, in which case the occurrence of an event would depend on the times of occurrence of the two preceding events. Given $\lambda(x,y)dt$, equal to the joint probability that an event has occurred in $(t, t+dt)$ with the two immediately preceding events at instants t_1, t_2, where $x = t - t_1$, $y = t_1 - t_2$, Wold derives integral equations for the distribution functions $F(x, y)$, which is the conditional probability that an event at t_1 is followed by an interval $(t_1, t_1 + x)$, given that the immediately preceding event occurred at $t_1 - y$, and $G(x)$, the absolute probability that an event at t_1 is followed by an event in $(t_1, t_1 + x)$. From these he obtains the probability $S(\nu, T)$ of exactly ν events in an arbitrary interval of length T. The definition of λ expresses the fact that the process is stationary, for λ remains the same if the set of events undergoes a translation along the time-axis.

This type of approach is adequate for one-dimensional fields because, by means of it, a process in which the course of events depends on all the prehistory up to a given time can be completely specified (apart from random variation); but it does not easily extend to more than one dimension. It seems in fact impossible to specify a stochastic process in two dimensions by means of a similar dependence of events. The difficulty is now (at least in the ecological application) that we are studying a function $F(x, y, T)$ say of the two space variables x, y and the time variable t, for a given value T of t, and we are given, specified by means of their space co-ordinates, a set of events which may have occurred separately at any time. Although the development of the process occurs along a time-axis, with events (i.e. new plants) occurring at specified times, it is mainly the consideration of the spatial pattern with which we are concerned.

For the purpose of finding probability relations between the numbers of plants in neighbouring finite areas, it is convenient to use the method of 'continuous parameter' stochastic processes, developed in connexion with physical problems such as electron cascades, the main contribution being jointly by Bhabha (1950) and Ramakrishnan (1950). They considered the case of particles with specific energies in a continuous range, i.e. effectively point processes in one dimension. Probability relations between the numbers of particles in non-overlapping infinitesimal ranges are defined by density functions of different orders, called by Ramakrishnan 'product densities', and the integrals of these over finite ranges yield linear functions of the product moments of the total numbers of particles in the finite ranges. The required mathematical treatment is obtained by considering a function of the continuous parameter E, $n(E)$ say, which equals 0 everywhere except at a finite number of points $E_1, E_2, ..., E_k$ say (at which events occur), where $n(E) = 1$. With suitable modifications this is the approach adopted in this paper, and in the next section the theory of spatial point processes is recapitulated in the most convenient manner. As in Wold, we assume that the processes are stationary in the statistical sense, the essential character being unaltered by any translations of the axes. This is a quite reasonable assumption for plant communities, where trends can often be assumed to be absent.

2. Specification of point stochastic processes in two dimensions

The distribution of plants over a region is described by a continuous parametric system, whose states are labelled by the two continuous variables x and y, the co-ordinates of position of a plant with respect to a given origin. Thus a given pair of values (x, y), which will alternatively be denoted by the parameter A, regarded as a vector, is taken to imply the occurrence of a plant in the infinitesimal rectangle with lower left-hand corner (x, y) and sides of length dx, dy. The infinitesimal element of area $dx\,dy$ will likewise be denoted by the vector dA. A realization of such a system will consist of a number of isolated points, k say, denoted by the parameter values $A_1, A_2, ..., A_k$. All states of the system are covered by allowing each A_i to range over the whole domain of A, and letting $k = 0, 1, 2, ...,$ to give all possible samples. With each possible sample is associated a probability $\Pi(A_1, A_2, ..., A_k)$, and a rigorous treatment of continuous parametric systems is obtained (see Bhabha, 1950) by a precise mathematical definition of the probability Π, using set theory. For the purposes of this paper a more direct approach, as given by Ramakrishnan (1950), is sufficient.

Let $N(A)$ denote the number of plants whose positions are below and to the left of the value $A \equiv (x, y)$, i.e. whose parametric values are less than A. Then we can consistently write

$$N(A) = \int^A dN(A),$$

and regard $dN(A)$ as denoting the number of plants in the infinitesimal rectangle, area dA, situated at A. Assume that the probability of one plant in dA is $m\,dA$, and the probability of more than one is $o(dA)$. Hence $dN(A)$ is a variable assumed to take only the values 0 or 1, and it follows that all moments of $dN(A)$ are equal to the probability that it takes the value 1. Also the consideration of probability relations between different $dN(A)$ is very much simplified, for a contribution to the product moment $E\{dN(A_1)...dN(A_k)\}$ will only occur when there is a plant in each dA_i, and the contribution is then simply unity. If any one of the dA_i contains no plant, the contribution is zero.

The probability relations between the numbers $dN(A)$ in the small areas dA are defined by functions of the form

$$f_k(A_1, A_2, ..., A_k)\,dA_1\,dA_2...dA_k = E\{dN(A_1)\,dN(A_2)...dN(A_k)\}, \tag{2·1}$$

which represents the joint probability that there is one plant in dA_1, one in dA_2, ..., one in dA_k, when $dA_1, dA_2, ..., dA_k$ are all separate non-overlapping areas. f_k is called a product density of degree k. $f_1(A)\,dA = E\{dN(A)\} = m\,dA$, and it may be noted that the integral of f_1 yields only the mean number of plants in the area of integration, as the addition rule of probabilities does not apply, the events in general not being mutually exclusive. When two of the dA_i are the same, a product density of degree k becomes one of degree $k-1$. For if $dA_k \equiv dA_{k-1}$,

$$E\{dN(A_1)\,dN(A_2)...dN(A_{k-1})^2\} = E\{dN(A_1)\,dN(A_2)...dN(A_{k-1})\}$$
$$= f_{k-1}(A_1, A_2, ..., A_{k-1})\,dA_1\,dA_2...dA_{k-1}.$$

It is degeneracies of this type that lead to the following theorem, given by Bhabha, for the product moment of the numbers $N(A_i') - N(A_i)$, denoted by $N_{(i)}$, in finite areas $[A_i, A_i']$ $(i = 1, 2, ..., k)$. A_i is the lower left-hand point, A_i' the upper right-hand point of the area,

which is assumed to be rectangular. The k areas denoted by the numbers $[1], [2], ..., [k]$, and any number of them may be the same area:

$$E\{N_{[1]}N_{[2]}...N_{[k]}\} = \int_{[1,...,k]} f_1(A)\,dA + \Sigma \int_{[1,...,r]}\int_{[r+1,...,k]} f_2(A_1, A_2)\,dA_1\,dA_2$$

$$+ \Sigma \int_{[1,...,r]}\int_{[r+1,...,s]}\int_{[s+1,...,k]} f_3(A_1, A_2, A_3)\,dA_1\,dA_2\,dA_3$$

$$+ ... + \int_{[1]}\int_{[2]}...\int_{[k]} f_k(A_1, ..., A_k)\,dA_1 ... dA_k, \tag{2·2}$$

where

(1) by $[1, ..., r]$ is meant the equality of the r areas $[1], ..., [r]$, if in fact they are equal. If they are not, the terms containing $[1, ..., r]$ do not appear in the expression;

(2) the summation sign before each term denotes a summation over like terms in which the integers $1, ..., k$ are distributed in all possible ways between the brackets affixed to the integrals, there being no distinction between the order of the integers in a bracket or between the order of the brackets. For example, with $k = 4$, integrals of the type of the second term are summed with the groupings of the integers given by

$$[1][234], \quad [2][341], \quad [3][412], \quad [4][123], \quad [12][34], \quad [13][24], \quad [14][23],$$

of the third term by

$$[1][2][34], \quad [1][3][24], \quad [1][4][23], \quad [2][3][14], \quad [2][4][13], \quad [3][4][12].$$

Thus, if the four areas are the same area $[1]$ say, we have for the expectation of the fourth power of the number in $[1]$

$$E\{N_{[1]}^4\} = \int_{[1]} f_1 + 7 \int_{[1]}\int_{[1]} f_2 + 6 \int_{[1]}\int_{[1]}\int_{[1]} f_3 + \int_{[1]}\int_{[1]}\int_{[1]}\int_{[1]} f_4,$$

using a condensed notation. If there is more than one area to consider, the expression is modified; for example,

$$E\{N_{[1]}^2 N_{[2]}^2\} = \int_{[1]}\int_{[2]} f_2 + \int_{[1]}\int_{[1]}\int_{[2]} f_3 + \int_{[1]}\int_{[2]}\int_{[2]} f_3 + \int_{[1]}\int_{[1]}\int_{[2]}\int_{[2]} f_4.$$

The product densities may consistently be regarded as factorial-moment densities (cf. Bartlett, 1954). For, if we consider the expression for $E\{N_{[1]}^r\}$, the coefficient of the term containing the product density of degree s is the sum of the coefficients of all terms

$$\Sigma ... \Sigma\, dN(A_1)^{\alpha_1}\, dN(A_2)^{\alpha_2} ... dN(A_s)^{\alpha_s}$$

for a given s and all different combinations of $\alpha_1, \alpha_2, ..., \alpha_s$ subject to the condition $\sum_{i=1}^{s} \alpha_i = r$.

That is, the coefficient of $\int_{[1]} ... \int_{[1]} f_s$ in $E\{N_{[1]}^r\}$ is $\Sigma r!/(\alpha_1! \alpha_2! ... \alpha_s!)$ the summation being over all possible different combinations of the $s\alpha_i$. Stevens (1937), in a different connexion, proves that this coefficient is of the general form $\Delta^s 0^r/s! = C_s^r$ say, where $\Delta^s 0^r$ $(s = 1, 2, ..., r)$ is the sth leading difference of the rth powers of the natural numbers. Thus we have

$$E\{N_{[1]}^r\} = C_1^r \int_{[1]} f_1(A)\,dA + C_2^r \int_{[1]}\int_{[1]} f_2(A_1, A_2)\,dA_1\,dA_2$$

$$+ ... + C_r^r \int_{[1]} ... \int_{[1]} f_r(A_1, ..., A_r)\,dA_1 ... dA_r.$$

One of Stevens's results is, in our notation,

$$N_{[1]}^r = C_1^r N_{[1]} + C_2^r N_{[1]}(N_{[1]} - 1) + \ldots + C_r^r N_{[1]}(N_{[1]} - 1) \ldots (N_{[1]} - r + 1).$$

A simple inductive argument then shows that

$$\int_{[1]} \ldots \int_{[1]} f_k(A_1, \ldots, A_k) \, dA_1 \ldots dA_k = E\{N_{[1]}(N_{[1]} - 1) \ldots (N_{[1]} - k + 1)\}.$$

For a process to be stationary we require that the covariance of the numbers in two areas is independent of their absolute position. Thus, for two areas [1], [2],

$$\mathrm{cov}\,(N_{[1]}, N_{[2]}) = E\{N_{[1]} - E(N_{[1]})\}\{N_{[2]} - E(N_{[2]})\}$$

$$= E\{N_{[1]} N_{[2]}\} - E\{N_{[1]}\} E\{N_{[2]}\}$$

$$= \int_{[1]} \int_{[1]} \{f_2(A_1, A_2) - m^2\} \, dA_1 \, dA_2, \quad \text{from (2.2).}$$

This implies that

$$f_2(A_1, A_2) - m^2 = \omega(A_2 - A_1),$$

so that $f_2(A_1, A_2)$ is a function of the differences $x_2 - x_1, y_2 - y_1$.

In practical applications the calculation of the product density of degree k will generally involve a tabulation of the different ways of getting a plant in each of the k infinitesimal areas, due usually to the presence of different classes of plants (different groups or families of plants, different generations, etc.). f_k is then obtained as the sum of the (mutually exclusive) probabilities for all these possible alternative cases, according to the addition rule of probabilities.

3. Analysis of variance on a grid of quadrats

The individual terms of the analysis of variance being considered in this paper are the sums of squares within blocks of a given size and between blocks of the next lowest size. The most useful type of grid to employ is one in which the number of quadrats is a power of 2, leading to blocks of $2, 4, 8, \ldots$ quadrats and consequently more information, relative to the size of the grid, than any other arrangement of blocks. The total size of the grid is quite arbitrary; for practical and theoretical purposes it has been fixed here at 256 quadrats, arranged in a 16×16 square. With a suitable choice of quadrat size, the important ecological effects should be found in the terms for the smaller-sized blocks, which have a fairly large number of degrees of freedom and are consequently subject to less fluctuation.

From the definition above, we see that if $B_{k(i)}$ denotes the total number of plants in the ith block of k quadrats, then S_k, the sum of squares within blocks of k quadrats and between blocks of $\frac{1}{2}k$ quadrats, is given by

$$S_k = \frac{2}{k} \sum_{i=1}^{512/k} B_{\frac{1}{2}k(i)}^2 - \frac{1}{k} \sum_{i=1}^{256/k} B_{k(i)}^2, \tag{3.1}$$

from which S_k has $n_k = 256/k$ degrees of freedom. This is ordinarily the most convenient practical way of calculating individual terms, but for deriving an expected analysis of variance and its sampling errors for any given theoretical model we need to work in terms

of the numbers in a single quadrat. Consider now the moments of S_k for a given k ($k = 2, 4, 8,$ ..., 256). A three-suffix notation for the quadrat numbers is the most convenient for this purpose.

Let N_{ijl} be the observed number of plants in quadrat (ijl) of the grid, where i ($= 1, 2, ...,$ 256 k) denotes a primary block of k quadrats, j ($= 1, 2$) a subblock of $\frac{1}{2}k$ quadrats, and l ($= 1, 2, ..., \frac{1}{2}k$) a quadrat of a subblock. We assume that blocks of the same size always have identical shapes. For $k = 2^{2p-2}$ ($p = 0, 1, 2, 3$) blocks are taken to be squares of sides 2^{p-1} quadrats, while for $k = 2^{2p-1}$ they are assumed rectangular, of dimensions $2^p \times 2^{p-1}$ quadrats, with the longer length always in the same direction.

From (3·1), transforming $B_{k(i)}$ to a sum of the N_{ijl}, we obtain

$$S_k = \frac{2}{k} \sum_i [(\sum_l N_{i1l})^2 + (\sum_l N_{i2l})^2 - \tfrac{1}{2}(\sum_j \sum_l N_{ijl})^2]$$

$$= \frac{1}{k} \sum_i [\sum_l (N_{i1l} - N_{i2l})]^2. \tag{3·2}$$

From (3·2), by long and tedious but otherwise simple algebra, the higher powers of S_k may be obtained. We quote only S_k^2 in full:

$$S_k^2 = \frac{1}{k^2} [\sum_i \{\sum_l (N_{i1l} - N_{i2l})\}^4 + \sum_{i \neq i'} \{\sum_l (N_{i1l} - N_{i2l})\}^2 \{\sum_l (N_{i'1l} - N_{i'2l})\}^2]$$

$$= \frac{1}{k^2} \sum_i \sum_j \{\sum_l N_{ijl}^4 + 4 \sum_{l \neq l'} N_{ijl}^3 N_{ijl'} - 4 \sum_{l, l'} N_{ijl}^3 N_{ijl'}$$

$$+ 3 \sum_{l \neq l'} N_{ijl}^2 N_{ijl'}^2 + 3 \sum_{l, l'} N_{ijl}^2 N_{ijl'}^2$$

$$+ 6 \sum_{l \neq l' \neq l''} N_{ijl}^2 N_{ijl'} N_{ijl''} + 6 \sum_{l' \neq l''} N_{ijl}^2 N_{ijl'} N_{ijl''}$$

$$- 12 \sum_{l \neq l'} N_{ijl}^2 N_{ijl'} N_{ijl''} + \sum_{l \neq ... \neq l''} N_{ijl} N_{ijl'} N_{ijl''} N_{ijl'''}$$

$$+ 3 \sum_{l \neq l', l'' \neq l'''} N_{ijl} N_{ijl'} N_{ijl''} N_{ijl'''} - 4 \sum_{l \neq l' \neq l''} N_{ijl} N_{ijl'} N_{ijl''} N_{ijl'''}\}$$

$$+ \frac{1}{k^2} \sum_{i \neq i'} \sum_{j, j'} \{\sum_{l, l'} N_{ijl}^2 N_{i'j'l'}^2 + 2 \sum_{l' \neq l''} N_{ijl}^2 N_{i'j'l'} N_{i'j'l''}$$

$$- 4 \sum_{l, l', l''} N_{ijl}^2 N_{i'j'l'} N_{i'j'l''} + \sum_{l \neq l', l'' \neq l'''} N_{ijl} N_{ijl'} N_{i'j'l''} N_{i'j'l'''}$$

$$+ 2 \sum_{l, l', l'', l'''} N_{ijl} N_{ijl'} N_{i'j'l''} N_{i'j'l'''} - 8 \sum_{l, l'; l'' \neq l'''} N_{ijl} N_{ijl'} N_{i'j'l''} N_{i'j'l'''}\}, \tag{3·3}$$

where $i, i'; j, j'; l, l', l'', l'''$ are varied independently over their whole respective ranges unless otherwise stated.

For the mean and variance of S_k, we take expected values of (3·2) and (3·3), and see that in the first case

$$E\{S_k\} = \frac{256}{k^2} [kE\{N_{ijl}^2\} + 2 \sum_{l \neq l'} E\{N_{ijl} N_{ijl'}\} - 2 \sum_{l, l'} E\{N_{ijl} N_{ijl'}\}], \tag{3·4}$$

and is only a linear combination of terms like $E\{N_{(1)}^2\}$, $E\{N_{(1)} N_{(2)}\}$, while $E\{S_k^2\}$ will be a linear combination of terms like $E\{N_{(1)}^4\}$, $E\{N_{(1)}^3 N_{(2)}\}$, $E\{N_{(1)}^2 N_{(2)}^2\}$, $E\{N_{(1)}^2 N_{(2)} N_{(3)}\}$, $E\{N_{(1)} N_{(2)} N_{(3)} N_{(4)}\}$.

The expressions for these expectations, obtained from (2·2), are given below, with $f_k(A_1, ..., A_k) dA_1 ... dA_k$ written f_k for short:

$$E\{N_{[1]}^2\} = \int_{[1]} f_1 + \int_{[1]} \int_{[1]} f_2, \quad E\{N_{[1]} N_{[2]}\} = \int_{[1]} \int_{[2]} f_2,$$

$$E\{N_{[1]}^4\} = \int_{[1]} f_1 + 7 \int_{[1]} \int_{[1]} f_2 + 6 \int_{[1]} \int_{[1]} \int_{[1]} f_3 + \int_{[1]} \int_{[1]} \int_{[1]} \int_{[1]} f_4,$$

$$E\{N_{[1]}^3 N_{[2]}\} = \int_{[1]} \int_{[2]} f_2 + 3 \int_{[1]} \int_{[1]} \int_{[2]} f_3 + \int_{[1]} \int_{[1]} \int_{[1]} \int_{[2]} f_4,$$

$$E\{N_{[1]}^2 N_{[2]}^2\} = \int_{[1]} \int_{[2]} f_2 + \int_{[1]} \int_{[1]} \int_{[2]} f_3 + \int_{[1]} \int_{[2]} \int_{[2]} f_3 + \int_{[1]} \int_{[1]} \int_{[2]} \int_{[2]} f_4,$$

$$E\{N_{[1]}^2 N_{[2]} N_{[3]}\} = \int_{[1]} \int_{[2]} \int_{[3]} f_3 + \int_{[1]} \int_{[1]} \int_{[2]} \int_{[3]} f_4,$$

$$E\{N_{[1]} N_{[2]} N_{[3]} N_{[4]}\} = \int_{[1]} \int_{[2]} \int_{[3]} \int_{[4]} f_4.$$

(3·5)

4. Application of the theory to an ecological model

The simplest assumption mathematically that can be made about a community of plants is that they are distributed at random in the Poisson distribution. Such a distribution might conceivably arise when an area is first invaded by wind-borne seeds, but it is hardly ever encountered in practice. A more realistic and more useful model is obtained by allowing plants distributed in this manner, i.e. at random, to become the centres of distribution ('parents') of a generation of 'offspring' plants, whose positions depend on those of their respective parents. The offspring of a given parent are assumed to be distributed independently of each other, the distances r from parent to offspring following the isotropic normal distribution

$$f(r) dr = e^{-\frac{1}{2}r^2/\sigma^2} dr/(2\pi\sigma^2),$$

(4·1)

and the numbers n of offspring per parent following an arbitrary probability law $p(n)$. We consider for simplicity the distribution of the offspring only, for then the only distances which have to be taken into account are those between offspring of the same group, and these are also isotropic normal, but with parameter $\sqrt{2}\,\sigma$. For, $f(r) dr$ is the product of two independent components $e^{-\frac{1}{2}x^2/\sigma^2} dx/\sqrt{(2\pi\sigma^2)}$, $e^{-\frac{1}{2}y^2/\sigma^2} dy/\sqrt{(2\pi\sigma^2)}$ along rectangular axes, and the distance between two offspring projected on either of these axes is merely the sum of two independent distances (from the parent) following a one-dimensional normal distribution.

The product density f_k of degree k is found (§ 2) as the sum of mutually exclusive probabilities, denoted by $\Pr(dA_1, ..., dA_k)$, for all the possible alternatives with k different areas $dA_1, ..., dA_k$. Here the contributions to f_k will arise from all different combinations of k plants from s groups ($s \leqslant k$) such that there are α_1 plants from the first group, ..., α_s from the sth with $\sum_{i=1}^{s} \alpha_i = k$. Of these we need only consider the case $s = 1$, i.e. all k plants from the same group; the other cases will merely result in products of lower order product densities since the contributions from different groups are entirely independent. We

prove now that the joint probability of k offspring of the same group occurring in $dA_1, ..., dA_k$ is

$$\Pr(dA_1, ..., dA_k) = m_0 E\{n(n-1)...(n-k+1)\} \phi(A_1, ..., A_k) dA_1 ... dA_k. \qquad (4\cdot2)$$

where

$$\phi(A_1, ..., A_k) = \left(\frac{1}{2\pi\sigma^2}\right)^{k-1} \frac{1}{k} \exp\left[-\sum_{i=1}^{k-1} \sum_{j>i} \frac{R_{ij}^2}{2k\sigma^2}\right], \left.\vphantom{\sum}\right\}$$

$$R_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2, \qquad (4\cdot3)$$

and m_0 is the mean density of the parent plants. The mean density of offspring plants is $m = m_0 E\{n\}$.

Assume that the position of the parent plant of a group with a random number n of offspring is at dA_0. The probability of this is

$$m_0 dA_0.$$

The probability of k ($\leqslant n$) offspring of this parent in $dA_1, ..., dA_k$, distant $R_{01}, ..., R_{0k}$ from dA_0, given the parents' position is at dA_0, is

$$n(n-1)...(n-k+1)\left(\frac{1}{2\pi\sigma^2}\right)^k \exp\left[-\sum_{i=1}^k \frac{R_{0i}^2}{2\sigma^2}\right] dA_1 ... dA_k,$$

since there are n possibilities for dA_1, $n-1$ for dA_2, ..., $n-k+1$ for dA_k, and each distance is distributed independently. Therefore, averaging over n,

$$\Pr(dA_0, dA_1, ..., dA_k)$$

$$= \Pr(dA_0) \Pr(dA_1, ..., dA_k \mid dA_0)$$

$$= m_0 E\{n(n-1)...(n-k+1)\}\left(\frac{1}{2\pi\sigma^2}\right)^k \exp\left[-\sum_{i=1}^k \frac{(x_0-x_i)^2 + (y_0-y_i)^2}{2\sigma^2}\right] dA_0 dA_1 ... dA_k$$

$$= m_0 E\{n(n-1)...(n-k+1)\} \frac{1}{2\pi\sigma^2} \exp\left[-k \frac{(x_0-\bar{x})^2 + (y_0-\bar{y})^2}{2\sigma^2}\right] dA_0$$

$$\times \left(\frac{1}{2\pi\sigma^2}\right)^{k-1} \exp\left[-\sum_{i=1}^{k-1} \sum_{j>i} \frac{(x_i-x_j)^2 + (y_i-y_j)^2}{2k\sigma^2}\right] dA_1 ... dA_k,$$

where

$$\bar{x} = \sum_{i=1}^k \frac{x_i}{k}, \quad \bar{y} = \sum_{i=1}^k \frac{y_i}{k}.$$

Integrating out with respect to $dA_0 (= dx_0 dy_0)$ over the whole region (assumed infinite), we obtain (4·2). $\phi(A_1, ..., A_k)$ may be shown to be a product of two $(k-1)$-variate normal distributions (in x and y), the $k-1$ variables being, for example, the differences of the first $k-1$ values of x (or y) from the kth, so that each variable is distributed normally, mean zero, variance $2\sigma^2$, and the correlation between any two variables is $\frac{1}{2}$.

For the mean values and variances of the set of sums of squares we require the product densities of second, third and fourth degree. With two areas dA_1 and dA_2, there are only two possibilities to consider: both areas might contain plants from different groups, in which case the joint probability $\Pr(dA_1, dA_2)$ is $m^2 dA_1 dA_2$, or else the plants may be offspring from the same group. We note that the case where one of the dA contains no plant is

automatically excluded, since it makes no contribution to the second degree product density. For the second case, we have from (4·2)

$$\Pr(dA_1, dA_2) = m_0 E\{n(n-1)\}\,\phi(A_1, A_2)\,dA_1\,dA_2,$$

so that, adding the two mutually exclusive probabilities, we obtain

$$f_2(A_1, A_2)\,dA_1\,dA_2 = [m_0 E\{n(n-1)\}\,\phi(A_1, A_2) + m^2]\,dA_1\,dA_2$$

$$= m\left(\frac{g_2}{4\pi\sigma^2}e^{-R_{12}^2/4\sigma^2} + m\right)dA_1\,dA_2, \tag{4·4}$$

from (4·3) where

$$g_r = E\{n(n-1)\dots(n-r+1)\}/E\{n\}. \tag{4·5}$$

Now tabulate the possibilities for three plants in three different areas dA_1, dA_2, dA_3. If the plants are all in different groups,

$$\Pr(dA_1, dA_2, dA_3) = m^3\,dA_1\,dA_2\,dA_3.$$

If dA_1, dA_2 are in a group G, dA_3 not in G, then since the contributions from different groups are independent, we have

$$\Pr(dA_1, dA_2, dA_3) = m_0 E\{n(n-1)\}\,\phi(A_1, A_2)\,dA_1\,dA_2 . m\,dA_3.$$

If dA_1, dA_2, dA_3 are all in the same group,

$$\Pr(dA_1, dA_2, dA_3) = m_0 E\{n(n-1)(n-2)\}\,\phi(A_1, A_2, A_3)\,dA_1\,dA_2\,dA_3.$$

This exhausts all the possibilities; therefore,

$$f_3(A_1, A_2, A_3) = m^3 + m^2 g_2\{\phi(A_1, A_2) + \phi(A_1, A_3) + \phi(A_2, A_3)\} + m g_3 \phi(A_1, A_2, A_3)$$

$$= m^3 + m^2 g_2 (e^{-R_{12}^2/4\sigma^2} + e^{-R_{13}^2/4\sigma^2} + e^{-R_{23}^2/4\sigma^2})/4\pi\sigma^2$$

$$+ m g_3 e^{-(R_{12}^2 + R_{13}^2 + R_{23}^2)/6\sigma^2}/12\pi^2\sigma^4. \tag{4·6}$$

We omit the tabulation for the fourth degree case, as it is obvious from the expression

$$f_4(A_1, A_2, A_3, A_4) = m^4 + m^3 g_2\{\phi(A_1, A_2) + \phi(A_1, A_3) + \phi(A_1, A_4) + \phi(A_2, A_3)$$

$$+ \phi(A_2, A_4) + \phi(A_3, A_4)\}$$

$$+ m^2 g_2^2\{\phi(A_1, A_2)\,\phi(A_3, A_4) + \phi(A_1, A_3)\,\phi(A_2, A_4)$$

$$+ \phi(A_1, A_4)\,\phi(A_2, A_3)\}$$

$$+ m^2 g_3\{\phi(A_1, A_2, A_3) + \phi(A_1, A_2, A_4) + \phi(A_1, A_3, A_4) + \phi(A_2, A_3, A_4)\}$$

$$+ m g_4 \phi(A_1, A_2, A_3, A_4). \tag{4·7}$$

The product densities are integrated and substituted in (3·5) to give the expectations needed in the calculation of $E\{S_k\}$, $E\{S_k^2\}$. An explicit solution has been found for the integral of f_2 only, but product densities of all degrees may be integrated numerically. Let the quadrats of the grid be squares of side h, and write

$$\int_{[1]}\int_{[2]}\dots\int_{[k]} \phi(A_1, \dots, A_k)\,dA_1\dots dA_k = h^2 I_{12\dots k}, \tag{4·8}$$

where square brackets round each suffix of I may be assumed if desired. Then equations (3·5) become

$$E\{N_{(1)}^2\} = mh^2 + mh^2 g_2 I_{11}, \quad E\{N_{(1)} N_{(2)}\} = mh^2 g_2 I_{12}.$$

$$\begin{aligned} E\{N_{(1)}^4\} = {} & mh^2 + 7m^2h^4 + 6m^3h^6 + m^4h^8 + mh^2(7 + 18mh^2 + 6m^2h^4) g_2 I_{11} \\ & + 3m^2h^4 g_2^2 I_{11}^2 + mh^2(6 + 4mh^2) g_3 I_{111} + mh^2 g_4 I_{1111}, \end{aligned}$$

$$\begin{aligned} E\{N_{(1)}^3 N_{(2)}\} = {} & m^2h^4 + 3m^3h^6 + m^4h^8 + 3m^2h^4(1 + mh^2) g_2 I_{11} + m^2h^4 g_3 I_{111} \\ & + mh^2(1 + 6mh^2 + 3m^2h^4) g_2 I_{12} + 3m^2h^4 g_2^2 I_{11} I_{12} \\ & + 3mh^2(1 + mh^2) g_3 I_{112} + mh^2 g_4 I_{1112}, \end{aligned}$$

$$\begin{aligned} E\{N_{(1)}^2 N_{(2)}^2\} = {} & m^2h^4 + 2m^3h^6 + m^4h^8 + 2m^2h^4(1 + mh^2) g_2 I_{11} \\ & + mh^2(1 + 4mh^2 + 4m^2h^4) g_2 I_{12} + m^2h^4 g_2^2(I_{11}^2 + 2I_{12}^2) \\ & + mh^2(1 + 2mh^2) g_3(I_{112} + I_{122}) + mh^2 g_4 I_{1122}, \end{aligned}$$

$$\begin{aligned} E\{N_{(1)}^2 N_{(2)} N_{(3)}\} = {} & m^3h^6 + m^4h^8 + m^3h^6 g_2 I_{11} + (m^2h^4 + 2m^3h^6) g_2(I_{12} + I_{13}) \\ & + (m^2h^4 + m^3h^6) g_2 I_{23} + m^2h^4 g_2^2(I_{11} I_{23} + 2I_{12} I_{13}) \\ & + (mh^2 + 2m^2h^4) g_3 I_{123} + m^2h^4 g_3(I_{112} + I_{113}) + mh^2 g_4 I_{1123}, \end{aligned}$$

$$\begin{aligned} E\{N_{(1)} N_{(2)} N_{(3)} N_{(4)}\} = {} & m^4h^8 + m^3h^6 g_2(I_{12} + I_{13} + I_{14} + I_{23} + I_{24} + I_{34}) \\ & + m^2h^4 g_2^2(I_{12} I_{34} + I_{13} I_{24} + I_{14} I_{23}) \\ & + m^2h^4 g_3(I_{123} + I_{124} + I_{134} + I_{234}) + mh^2 g_4 I_{1234}. \end{aligned}$$

$$(4·9)$$

The details of the calculations are omitted, as much tedious manipulation is involved.

Table 1. *Expected mean squares for the non-random model*

k	2	4	8	16	32	64	128	256
n_k	128	64	32	16	8	4	2	1
$E\{S_k'/mh^2\}$	1·12	1·19	1·64	1·84	2·40	2·58	2·91	3·02

There are many different numerical examples possible, since there are two sets of parameters to be varied, those of $p(n)$ and those of $f(r)$. A study of the effects of different functions $p(n)$ and $f(r)$ on the form of the distribution of $N_{(i)}$, the number of plants in a single quadrat, has already been made (Thompson, 1954). Table 1 gives the expected mean squares $E\{S_k'/mh^2\}$, $(S_k' = S_k/n_k)$, for the particular case $\sigma = h/\sqrt{2}$ (which gives a simplification of the numerical integrations) and a binomial distribution for $p(n)$ with a mean of three offspring per group. The parameters in $p(n) = \binom{N}{n} p^n (1-p)^{N-n}$ $(n = 0, 1, ..., N)$ are $N = 6$, $p = \frac{1}{2}$, and $g_r = (N-1)(N-2) ... (N-r+1) p^{r-1}$. The set of expected mean squares $E\{S_k'\}$ has upper and lower limits, reached as σ tends to 0 and ∞ respectively; they are

$mh^2(1+g_2)$ and mh^2, independent of k. These are both mean squares for a Poisson distribution, and are to be expected, since in the first case each offspring's position coincides with that of its parent, and in the second case there is effectively no clumping.

The standard error of a sum of squares is not in itself a very useful statistic, because of the skew distribution of S_k, at least for a small number of degrees of freedom. For the first three mean squares of Table 1, with fairly large n_k (and for which normal approximations might be expected to hold), the standard errors are, respectively, 0·20, 0·27, 0·46 for $mh^2 = 1$, and 0·13, 0·16, 0·26 for $mh^2 = \frac{1}{2}$. A more useful practical method of studying the sampling variation is to determine the limits of error of individual mean squares, so that the set of expected mean squares for a given model has appropriate to it 'significance bands' outside which terms observed from samples of the model would only be expected to fall with a given (small) probability. In this approach we ignore the fact that a real correlation exists between any pair of mean squares, except for the case when the $N_{[i]}$ are independent normal variables, and calculate limits of error on the assumption of approximate χ^2 distributions for individual mean squares, based on their first two moments, since we know that for $N_{[i]}$ normal (variance σ^2 say) S_k/σ^2 is distributed exactly as χ^2 with n_k degrees of freedom. Table 2 gives these limits of error for the case $mh^2 = 1$ only, but to a first approximation they may be taken to apply to mean squares $S_k'/(mh^2)$ so long as mh^2 is not too different from unity. The bands are symmetrical, in the sense that the probability of an observed mean square being above the upper limit equals the probability of its falling below the lower limit.

Table 2. *Approximate p % significance bands for the mean squares of Table 1 ($mh^2 = 1$)*

k	2	4	8	16	32	64	128	256
95 % band	0·76	0·72	0·86	0·72	0·62	0·28	0·06	0·00
	1·55	1·78	2·65	3·50	5·38	7·28	10·83	15·22
80 % band	0·86	0·86	1·07	0·99	0·96	0·62	0·26	0·03
	1·39	1·55	2·25	2·78	4·08	5·08	6·72	8·18

It is instructive to compare these figures (applying to a model with not a very high degree of contagion) with the results for a purely random distribution (parents alone, say), which may be shown to approximate very closely to the normal case. The correlations between individual mean squares are in this random case almost negligible, and S_k/mh^2 follows almost exactly a χ^2 distribution with $n_k\{2kmh^2/(1+2kmh^2)\}$ degrees of freedom. For a Poisson distribution, we have explicitly

$$E\{S_k'\} = mh^2,$$
$$\mathrm{var}\{S_k'\} = mh^2(1+2kmh^2)/256. \qquad (4.10)$$

Table 3 gives the limits of error for the case $mh^2 = 1$, but as in Table 2 they may equally well apply for other values of mh^2. In fact, for $mh = \frac{1}{2}, \infty$, which are symmetrically placed with respect to $mh^2 = 1$ here, the differences from the values given in the tables are never more than ± 0.02.

Table 3. *Approximate p % significance bands for the mean squares*
from a Poisson distribution $(mh^2 = 1)$

k	2	4	8	16	32	64	128	256
95 % band	0·75	0·67	0·56	0·42	0·27	0·12	0·03	0·00
	1·29	1·40	1·56	1·82	2·20	2·79	3·70	5·03
80 % band	0·83	0·77	0·69	0·58	0·43	0·26	0·10	0·02
	1·18	1·25	1·34	1·48	1·68	1·95	2·30	2·71

5. DISCUSSION

When an analysis of variance is carried out on a set of observational data, the usual aim is to test the homogeneity of the set of variances obtained, employing the F test. With a grid of quadrats, the appropriate method would be to test the ratio $(S_k/n_k) \cdot (S_2/n_2)$, which, on the null hypothesis that the numbers $N_{(i)}$ in the quadrats are independently and normally distributed with the same variance, follows the F distribution with n_2 and n_k degrees of freedom. The conditions for the applicability of this test will not, however, generally hold in the present ecological application, for the associations of plants in clumps result in the non-independence of the numbers in neighbouring quadrats, and even more distant quadrats if the clumps themselves are related, and the effect of clumping is usually to produce a contagious distribution of the quadrat numbers. Under the usual null hypothesis, H_0 say, we have

$$\Pr\{F \geqslant F_\alpha \mid H_0\} = \alpha, \tag{5·1}$$

where $F = (S_k/n_k)/(S_2/n_2)$, and F_α is the value of F on n_2 and n_k degrees of freedom at the 100α % level of significance. The expected value of F on this hypothesis is unity; however, for most non-random models, including the example above, it is greater. (It may occasionally be less than unity.) Therefore, to study the power function of the F test we should ideally calculate the probability

$$\Pr\{F/E(F) \geqslant F_\alpha \mid H\} = \beta, \tag{5·2}$$

where H is the non-normal hypothesis relating to the particular model, and compare β with α.

This has been done approximately for the example of § 4, for the two cases $k = 4, 8$ and with $\alpha = 0·05, 0·01$, the results being given in Table 4. The method used is one recommended by David & Johnson (1951). Rewrite (5·2) in the form

$$\Pr\{S_k - aS_2 \geqslant 0 \mid H\} = \beta, \tag{5·3}$$

where $a = n_k F_\alpha E(F)/n_2 = 2F_\alpha E(F)/k$. The first two moments of $S = S_k - aS_2$ are sufficient to find this probability approximately, transforming S so that probability levels for the χ^2 distribution can be used. A new statistic $S/(mh^2) + C$ is formed whose first two moments are exactly those of a χ^2 distribution with f degrees of freedom by taking

$$C = \tfrac{1}{2}\mathrm{var}\{S/(mh^2)\} - E\{S/(mh^2)\}, \quad f = \tfrac{1}{2}\mathrm{var}\{S/(mh^2)\}.$$

Then $\Pr\{\chi^2_{(f)} \geqslant C\}$ furnishes β. The method is highly satisfactory when the $N_{(i)}$ are Poisson or normal variables. These results show that the actual level of significance being used is

Table 4. *Values of β (from (5·3)) for the model of §4*

k $\diagdown$ mh^2	$\alpha = 0.05$			$\alpha = 0.01$		
	$\frac{1}{2}$	1	∞	$\frac{1}{4}$	1	∞
4	0·090	0·073	0·055	0·033	0·023	0·013
8	0·117	0·103	0·089	0·048	0·038	0·028

quite close to its assumed true value. However, we have assumed on *a priori* knowledge of $E\{F\}$, which would not normally be so in practice. A truer picture of what would happen if the F test were applied indiscriminately is given in Table 5, where we have calculated

$$\Pr\{F \geqslant F_\alpha \mid H\} = \beta'. \tag{5·4}$$

In this particular example, we see that the use of the F test on unadjusted variances is of no value at all for k greater than 4, and even for $k = 4$ the differences between β' and α are becoming appreciable.

Table 5. *Table of β' (from (5·4)), for the model of §4*

k $\diagdown$ mh^2	$\alpha = 0.05$			$\alpha = 0.01$		
	$\frac{1}{2}$	1	∞	$\frac{1}{4}$	1	∞
4	0·139	0·119	0·098	0·054	0·040	0·027
8	0·535	0·537	0·539	0·350	0·342	0·334

We have only discussed one particular example in this paper. Several others giving rise to non-random distributions of plants have been derived, consistent with the mathematical and computational difficulty involved. They are described at length in Thompson (1955), with more emphasis on the ecological application, however, so that it may be of interest to note some of them briefly here. We can extend the model of §4 to include several generations of plants, the distances from each offspring to its parent still following the isotropic normal distribution (4·1). The offspring of the original parent become the parents of another generation of offspring, which in their turn become parent plants, and so on. The distance between any two plants of the same group also follows (4·1), but with parameter depending on the number of direct steps between the two plants. If the original parents are randomly distributed, we find similarly to (4·4),

$$f_2(A_1, A_2) - m^2 = m_0 \sum_\alpha E\{2n_\alpha\} e^{-R^2/2\alpha\sigma^2}/2\pi\alpha\sigma^2. \tag{5·5}$$

where n_α is the number of pairs of plants in the group which have the distances between them distributed with parameter $\sqrt{\alpha}\,\sigma$.

Models using the isotropic normal distribution for $f(r)$ produce over-dispersion of individuals (ratio of variance to mean in a single quadrat greater than unity). If the x and y components of the distance between successive plants in a chain follow independent χ^2 distributions with $2f$ degrees of freedom ($f > 2$), then under-dispersion of individuals results because of the 'negative correlation' effect introduced (cf. Bartlett, 1954). The frequency function of the distance between plants g generations apart is easily obtained, being simply the product of two independent χ^2 distributions with $2fg$ degrees of freedom. For a group with n plants (the nth being the offspring of the $(n-1)$th, and so on), the original parent being randomly distributed and with $f = 2$, we have

$$f_2(A_1, A_2) - m^2 = m_0 \sum_{g=1}^{n-1} 2(n-g)(\lambda_1 \lambda_2)^{2g} [(x_2 - x_1)(y_2 - y_1)]^{2g-1} e^{-\lambda_1(x_2-x_1)-\lambda_2(y_2-y_1)} [(2g-1)!]^2,$$

$$(5 \cdot 6)$$

where λ_1, λ_2 are arbitrary and are considered in relation to the quadrat size h, and $x_2 \geqslant x_1 \geqslant 0$. $y_2 \geqslant y_1 \geqslant 0$. This represents approximately an actual field example, in which a definite forward move is made in each generation due to vegetative spreading of the plants, and the probability of having two successive plants very close together is small. The other models discussed are developments of these two basic types and have no intrinsic mathematical interest. In all of them, however, an attempt has been made to keep the ecological aspect in mind so that they should be descriptive in some way of an idealized plant community.

REFERENCES

Bartlett, M. S. (1954). Processus stochastiques ponctuels. *Ann. Inst. H. Poincaré*, **14**, 35.

Bhabha, H. J. (1950). On the stochastic theory of continuous parametric systems and its application to electron cascades. *Proc. Roy. Soc. A*, **202**, 301.

David, F. & Johnson, N. L. (1951). The effect of non-normality on power function of the F test in the analysis of variance. *Biometrika*, **38**, 43.

Greig-Smith, P. (1952). The use of random and contiguous quadrats in the study of the structure of plant communities. *Ann. Bot.*, N.S., **16**, 293.

Ramakrishnan, A. (1950). Stochastic processes relating to particles distributed in a continuous infinity of states. *Proc. Camb. Phil. Soc.* **46**, 595.

Stevens, W. L. (1937). Significance of grouping. *Ann. Eugen., Lond.*, **8**, 57.

Thompson, H. R. (1954). A note on contagious distributions. *Biometrika*, **41**, 268.

Thompson, H. R. (1955). Statistical study of plant distribution patterns using a grid of contiguous quadrats (in the Press).

Wold, H. (1949). Sur les processus stationnaires ponctuels. Le calcul des probabilités et des applications. *Publications du C.N.R.S.*, **13**.

THE OUTCOME OF A STOCHASTIC EPIDEMIC—A NOTE ON BAILEY'S PAPER

By P. WHITTLE

Applied Mathematics Laboratory, New Zealand Department
of Scientific and Industrial Research

In a recent paper (1953) N. Bailey has considered a stochastic epidemic model of the type set up by Bartlett (1949), and has shown that the probability distribution (P_w) of the ultimate number of infected individuals (w) may be calculated by solving a certain set of doubly recurrent relations. I propose to show that for quite a general case these same probabilities may be obtained by the solution of a set of singly recurrent relations (eqs. (17) and (24)). Furthermore, an expression may be derived (eqs. (38) and (40)) for the probability that an infection introduced into a large population will 'take'—this provides a stochastic equivalent to Kermack & McKendrick's threshold theorem (1927 and later).

1. INTRODUCTORY

Following Bailey, we shall assume that an initial a infectious cases are introduced into a population of n uninfected but susceptible individuals, and that the probability that after a time t there are r susceptibles still uninfected and s infectious cases not yet removed is $p_{rs}(t)$. Let $Bs\Delta t$ be the probability that an infectious individual is removed in the infinitesimal time interval $(t, t+\Delta t)$, and let $A_r s\Delta t$ be the corresponding probability that a new infection takes place. No particular form is assumed for the function A_r at the moment.

The development of the probabilities p_{rs} is then governed by the relations

$$\frac{\partial}{\partial t} p_{rs} = A_{r+1}(s-1) p_{r+1,s-1} + B(s+1) p_{r,s+1} - (A_r s + Bs) p_{rs}. \tag{1}$$

which become

$$(\lambda + A_r s + Bs) q_{rs} = A_{r+1}(s-1) q_{r+1,s-1} + B(s+1) q_{r,s+1} + \delta_{nr}\delta_{as} \tag{2}$$

if the transformation

$$q_{rs} = \int_0^\infty e^{-\lambda t} p_{rs}(t)\, dt \tag{3}$$

is performed. As Bailey observes, the probability of an epidemic of total size w is then

$$P_w = B f_{n-w,1}, \tag{4}$$

where

$$f_{rs} = \lim_{\lambda \to 0} q_{rs}. \tag{5}$$

This limit exists as long as $s > 0$, and the equations regulating the appropriate f_{rs} are obtained simply by setting λ equal to zero in equation (2), with $s = 1, 2, \ldots$. These are, in effect, the equations used by Bailey for the computation of the P_w.

2. ESTABLISHMENT OF THE RECURRENCE RELATIONS

If we write

$$h_{r0} = 0, \quad h_{rs} = s f_{rs} \quad (s = 1, 2, \ldots), \tag{6}$$

$$\alpha_r = \frac{B}{A_r + B}, \tag{7}$$

$$\beta_r = \frac{A_{r+1}}{A_r + B}, \tag{8}$$

then the equations for the f_{rs} take the form

$$h_{rs} = \alpha_r h_{r,s+1} + \beta_r h_{r+1,s-1},$$ (9)

$$h_{ns} = \alpha_n h_{n,s-1} + \beta_n \left(\frac{\delta_{as}}{A_{n+1}}\right) \quad (r = n-1, n-2, \ldots; s = 1, 2, \ldots).$$ (10)

Now, let

$$H_r(x) = \sum_{s=1}^{\infty} h_{rs} x^{s+1}.$$ (11)

Multiplying (9) by x^{s+1} and summing over s we find

$$H_r(x) = \frac{x^2}{x - \alpha_r} [\beta_r H_{r+1}(x) - \alpha_r h_{r1}].$$ (12)

A direct solution of (9) shows that

$$h_{r1} = \frac{\beta_r}{\alpha_r} H_{r+1}(\alpha_r),$$ (13)

as, indeed, it must, if the expression (12) for H_r is to constitute a finite series in x. We have thus

$$H_r(x) = \frac{\beta_r x^2}{x - \alpha_r} [H_{r+1}(x) - H_{r+1}(\alpha_r)],$$ (14)

a relation which certainly holds for $r = n-1, n-2, \ldots$, and also for $r = n$ if we introduce a function

$$H_{n+1}(x) = \frac{x^a}{A_{n+1}}$$ (15)

(cf. eq. (10)). Further, by (4), (6) and (13) we have

$$P_w = B \frac{\beta_{n-w}}{\alpha_{n-w}} H_{n-w+1}(\alpha_{n-w}).$$ (16)

The required probabilities P_w can thus be derived by solving the simply recurrent relation (14) with initial condition (15), and then using (16). By doing almost precisely this we shall obtain an explicit recurrence relation for the P_w. With the help of (14) and (15) $H_{n-w+1}(x)$ can be expressed in terms of $H_{n-w+2}(\alpha_{n-w+1}) \ldots H_n(\alpha_{n-1}), H_{n+1}(x)$. Setting $x = \alpha_{n-w}$ in this expression and substituting for the $H_{r+1}(\alpha_r)$ from (16) we find

$$\sum_{w=0}^{u} \kappa_{n-u+1,n-u} \kappa_{n-u+2,n-u} \cdots \kappa_{n-w,n-u} \frac{\alpha_{n-w}}{\beta_{n-w}} P_w = \kappa_{n-u+1,n-u} \kappa_{n-u+2,n-u} \cdots \kappa_{n,n-u} \left(\frac{B}{A_{n+1}}\right) \alpha_{n-u}^a,$$ (17)

where

$$\kappa_{rr} = 1, \quad \kappa_{rs} = \frac{\alpha_s^2 \beta_r}{\alpha_s - \alpha_r} \quad (r \neq s).$$ (18)

The final relation of (17), for $u = n$, reduces to

$$\sum_0^n P_w = 1,$$ (19)

provided that A_0 is zero, as it must be.

Complications arise if any of the A_r's (and consequently the corresponding α_r's) are equal. When this happens P_w involves not only $H_{n+1}(x)$ but also its derivatives. The extreme case in this direction is that for which A_r is constant for $r > 0$

$$A_r = A \quad (r = 1, 2, \ldots),$$
$$A_0 = 0,$$ (20)

as is approximately the case during the initial stages of an epidemic if the number of susceptibles is large. The appropriate modification of (17) may be derived by a repeated application of de l'Hôpital's rule, and may be shown by induction to have a solution

$$
\left.\begin{array}{l}
P_w = \dfrac{A^w B^{a+w}}{(A+B)^{a+2w}} \dfrac{a(a+2w-1)!}{w!\,(a+w)!} \\[2mm]
P_n = 1 - \sum_0^{n-1} P_w
\end{array}\right\} \quad (w = 0, 1, ..., n-1). \tag{21}
$$

(Formula (21) is a generalization of one obtained by D. G. Kendall in a slightly different context, see formula (52) of his paper of 1948.) We shall return to this case later, and shall for the moment consider the more usual alternative

$$
A_r = Cr, \tag{22}
$$

where C is a constant. We have

$$
\left.\begin{array}{l}
\alpha_r = \dfrac{B}{B+Cr}, \qquad \beta_r = \dfrac{C(r+1)}{B+Cr}, \\[2mm]
\kappa_{rs} = \dfrac{B(r+1)}{(B+Cs)(r-s)} = \dfrac{\alpha_s(r+1)}{r-s},
\end{array}\right\} \tag{23}
$$

and (17) becomes

$$
\sum_{w=0}^{u} \binom{n-w}{n-u} \alpha_{n-u}^{-w} P_w = \binom{n}{u} \alpha_{n-u}^{a} \quad (u = 0, 1, ..., n). \tag{24}
$$

For computational purposes it is convenient to consider instead of P_w the quantity

$$
Q_w = \frac{(n-w)!}{n!} P_w, \tag{25}
$$

for which

$$
\sum_{w=0}^{n} \frac{\alpha_{n-u}^{-w} Q_w}{(u-w)!} = \frac{\alpha_{n-u}^{a}}{u!}. \tag{26}
$$

3. THE PROBABILITY OF EPIDEMIC

Let us now return to expression (21). We shall use this as a comparison formula for establishing the behaviour of more refined models. Note, however, that the model upon which it is based is a perfectly valid one, which for quite large ranges of w is more realistic than that corresponding to assumption (22), since this assumption requires that the population mix homogeneously, a requirement never fulfilled in a large population.

The following table gives the first few values of P_w as calculated from formulae (21) and (24) for the case $a = 2$, $n = 30$, $\rho = B/C = 30B/A = 10$:

	w	0	1	2	3
P_w	$A_r = 3$	0·0625	0·0234	0·0110	0·0058
	$A_r = 0·1r$	0·0625	0·0251	0·0122	0·0078

The difference between the two sets of probabilities is small but increases with increasing w.

We shall now adopt the following definition: It shall be said that an epidemic has (has not) taken place if the total proportion of susceptibles which become infected exceeds (does not exceed) a predetermined fraction γ. With this definition the probability of no epidemic is

$$\pi_\gamma = \sum_{w=0}^{n\gamma} P_w, \tag{27}$$

where P_w is in general given by (17).

Consider now two other models for which the infection intensities are given by A_r', A_r''. We shall assume that in all three cases the infection intensity is non-decreasing with increasing r:

$$\left.\begin{aligned} A_{r+1} &\geqslant A_r \\ A_{r+1}' &\geqslant A_r' \\ A_{r+1}'' &\geqslant A_r'' \end{aligned}\right\} \quad (r = 0, 1, \ldots). \tag{28}$$

We shall further assume that, at least for the range $n \geqslant r \geqslant n(1-\gamma)$, the intensity for the first model lies uniformly between those for the other two

$$A_r' \geqslant A_r \geqslant A_r''. \tag{29}$$

It is then intuitively evident that

$$\sum_0^{n\gamma} P_w' \leqslant \sum_0^{n\gamma} P_w \leqslant \sum_0^{n\gamma} P_w''. \tag{30}$$

Suppose now that the intensities for those two comparison models have the constant values

$$\left.\begin{aligned} A_r' &= A_n \\ A_r'' &= A_{n(1-\gamma)} \end{aligned}\right\} \quad (r > 0), \tag{31}$$

while $A_0' = A_0'' = 0$. Condition (29) is thus fulfilled, and the inequality (30) becomes

$$\sum_0^{n\gamma} S_w(A_n) \leqslant \pi_\gamma \leqslant \sum_0^{n\gamma} S_w(A_{n(1-\gamma)}), \tag{32}$$

where we have used $S_w(A)$ to denote the expression in (21).

Consider now the partial sum $\sum_0^{n\gamma} S_w(A)$ as n becomes large. We have

$$\frac{S_{w+1}}{S_w} = \frac{AB}{(A+B)^2} \frac{(a+2w+1)(a+2w)}{(w+1)(a+w+1)} < \frac{4AB}{(A+B)^2} \quad \text{if} \quad w > \frac{(a+1)(a-4)}{6}$$
$$= 4k \quad \text{(say)}. \tag{33}$$

The quantity $4k$ will be less than unity, except for the case $A = B$, which we shall exclude for the moment. We can thus write

$$\sum_0^{n\gamma} S_w(A) = \sum_0^\infty S_w(A) - R_{n\gamma}(A), \tag{34}$$

where $\qquad R_{n\gamma}(A) = \sum_{n\gamma+1}^\infty S_w(A) < S_{n\gamma+1}(A)[1 + (4k) + (4k)^2 + \ldots] = O[(4k)^{n\gamma}]. \tag{35}$

The infinite sum in (34) has the value

$$\sum_0^\infty S_w(A) = \left[\frac{1 - \sqrt{(1-4k)}}{2k}\right]^a \left(\frac{B}{A+B}\right)^a = \left[\frac{A+B-|A-B|}{2A}\right]^a. \tag{36}$$

Combining (32). (34). (35) and (36) we have thus

$$\left[\frac{A_n + B - |A_n - B|}{2A_n}\right]^a + O[(4k)^{n\gamma}] \leqslant \pi_\gamma \leqslant \left[\frac{A_{n(1-\gamma)} + B - |A_{n(1-\gamma)} - B|}{2A_{n(1-\gamma)}}\right]^a. \tag{37}$$

For large n the remainder term in the first member of inequality (37) is quite negligible, and we shall no longer include it. Note now from (36) that we have $\Sigma S_w(A) = 1$ or $(B/A)^a$ according as A is less or greater than B. Setting these evaluations in (37) we find that there are at least three distinct cases:

$$\left.\begin{aligned}
A_n > B, \; A_{n(1-\gamma)} > B: \quad & (B/A_n)^a \leqslant \pi_\gamma \leqslant (B/A_{n(1-\gamma)})^a, \\
A_n > B, \; A_{n(1-\gamma)} < B: \quad & (B/A_n)^a \leqslant \pi_\gamma \leqslant 1, \\
A_n < B, \; A_{n(1-\gamma)} < B: \quad & \pi_\gamma = 1.
\end{aligned}\right\} \tag{38}$$

We may sum up the situation in terms of Bailey's removal ratio

$$\rho_n = B/C = nB/A_n, \tag{39}$$

the ratio of removal and infection rates for a population of size n:

> For $\rho_n < n$ and $\rho_{n(1-\gamma)} < n(1-\gamma)$, the probability of epidemic lies between
>
> $1 - \left(\dfrac{\rho_n}{n}\right)^a$ and $1 - \left(\dfrac{\rho_{n(1-\gamma)}}{n(1-\gamma)}\right)^a$,
>
> For $\rho_n < n$ and $\rho_{n(1-\gamma)} > n(1-\gamma)$, the probability of epidemic lies between
>
> zero and $1 - \left(\dfrac{\rho_n}{n}\right)^a$,
>
> For $\rho_n > n$ and $\rho_{n(1-\gamma)} > n(1-\gamma)$, the probability of epidemic is zero.

$$\tag{40}$$

These statements provide an equivalent in the stochastic case to Kermack & McKendrick's threshold theorem in the deterministic case, at least for the case of large populations. Since for large n the ratio $\rho_n/n = B/A_n$ will tend to an almost constant value, statements (40) may be roughly condensed to

> For $\rho_n < n$ the probability of epidemic is $1 - (\rho_n/n)^a$,
>
> For $\rho_n > n$ the probability of epidemic is zero.

$$\left.\right\} \tag{41}$$

The transition case $\rho_n = n$ cannot be adequately treated by approximate considerations of the present type. Equation (28) indicates that the probability of completion is of fairly constant magnitude for small w, roughly of order $(\frac{1}{2})^a$. As the number of susceptibles diminish, however, the critical value of ρ will fall, and it seems likely that the epidemic will eventually be halted, although only after having made appreciable inroads on the population.

The statements (36) are reminiscent of similar statements concerning the natural extinction of populations (cf. Bartlett, 1946) and could have been derived by regarding the group of infected persons as a population with birth and death rates A/n and B. To reason in this way is unsatisfactory, however, since the condition that the infected group shall ultimately disappear provides no guarantee that infection will be confined to a preassigned fraction of the population of susceptibles.

The argument above is probably more illuminating in the following intuitive form. The probability distribution P_w of (24) presents two different forms according as A is less than

or greater than B (see Fig. 1). In both cases P_w dwindles with increasing w, and if the population size is large enough to permit w to take large values P_w will finally approach zero. In the case $A < B$ the sum of the P_w up to this stage will approach unity. In the case $A > B$ this sum will have some value less than unity, say $1 - \alpha$, so that P_n must have a finite value α if relation (19) is to be fulfilled.
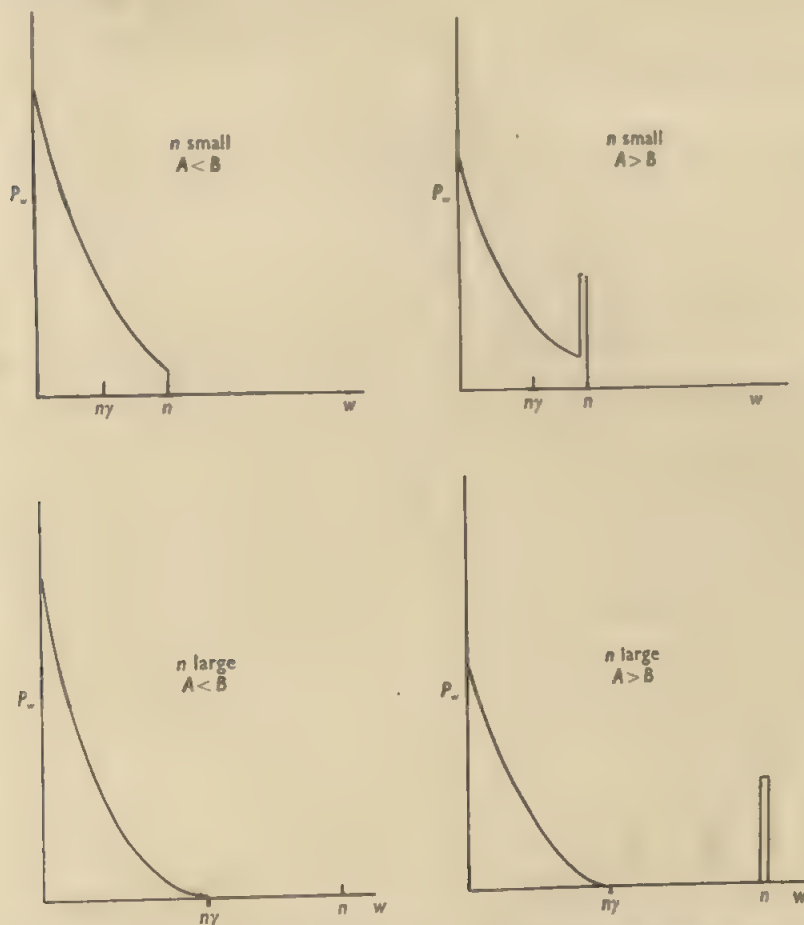
Fig. 1. The appearance of the distribution curve P_w of the ultimate number of infected persons, on the assumption of a constant infection rate.

For large n the probability of no epidemic

$$\pi_\gamma = \sum_0^{n\gamma} S_w(A) \tag{42}$$

will be equal to the area under the initial part of the curve: unity if $A < B$, α if $A > B$.

The fact that all probability mass which does not fall in the first J-shaped part of the curve falls at $w = n$ indicates that either the epidemic keeps within bounds (probability $1 - \alpha$) or else it infects the entire population (probability α).

Models with varying A show a similar, although less extreme behaviour. Thus the distribution curves calculated by Bailey are either J-shaped or U-shaped, depending upon the relative values of the removal ratio and the population size. The J-shaped curves

correspond to cases in which the infection is almost certainly confined to a small proportion of the population. The U-shaped curves correspond to cases in which the infection strikes at *either* a small proportion *or* a large proportion of the population, the probabilities of these two alternatives being equal to the integrals of the corresponding limbs of the probability distribution.

What our argument asserts in effect is that for a large population the form and integral of the first limb of the distribution P_w is equal to that calculated on the assumption of a constant infection intensity.

REFERENCES

BAILEY, N. T. J. (1953). The total size of a general stochastic epidemic. *Biometrika*, **40**, 177.

BARTLETT, M. S. (1946). Stochastic Processes. Mimeographed, North Carolina lecture notes.

BARTLETT, M. S. (1949). Some evolutionary stochastic processes. *J. R. Statist. Soc.* B, **11**, 211.

KENDALL, D. G. (1948). On the generalized 'birth-and-death' process. *Ann. Math. Statist.* **19**, 1.

KERMACK, W. O. & McKENDRICK, A. G. (1927 and later). Contributions to the mathematical theory of epidemics. *Proc. Roy. Soc.* A, **115**, 700; **138**, 55; **141**, 94.

A NOTE ON BAILEY'S AND WHITTLE'S TREATMENT OF A GENERAL STOCHASTIC EPIDEMIC

By F. G. FOSTER

Division of Research Techniques, London School of Economics

1. *Introduction.* In a recent paper Bailey (1953) obtained a set of doubly recurrent relations for the probability distribution (P_w) of the total size (w) of a stochastic epidemic. Bailey also quotes an explicit formula for P_w due to the author but observes that this is not suitable for computation. In a note on Bailey's paper, Whittle (1955)* showed that the P_w could be more simply calculated from a set of singly recurrent relations.

In the present note, Whittle's relations are considered. The notation is simplified by use of symmetric functions, and it is shown how a set of singly recurrent relations may be obtained for P_w in a quite general case by use of a simple probability argument. Whittle's relations may then be re-derived as a special case of these relations.

Following Bailey and Whittle, we consider a population which consists initially of n uninfected but susceptible individuals and a infected individuals. At any time t, if there are r susceptibles and s infected cases, the probability of one new infection taking place in time dt is $rs\,dt$ and the probability of one infected being removed from circulation is $\rho s\,dt$. The epidemic ends whenever either all infected have been removed or the whole population of $a + n$ individuals has become infected. If when the epidemic ends the number of new infections is w (i.e. not counting the original a infections), we say that the epidemic is of total size w, and we denote by P_w the probability of this event.

2. The problem may be treated as a random walk if we fix attention on the fluctuating number of *newly* infected, ξ, present at any time. At any instant, ξ may be either increased by unity or decreased by unity, and we consider the sequence of instants at which such changes occur.

Let us now represent the behaviour of ξ by the motion of a particle on a rectangular lattice, such that a move from (x, y) to $(x + 1, y)$ represents a unit increase in ξ, and a move from (x, y) to $(x, y + 1)$ represents a unit decrease. Thus, at any instant, the x-co-ordinate will represent the total number of infected cases (in addition to the initial a) and the y-co-ordinate the total number of removals which has occurred up to that instant. We suppose that the probabilities of the moves from (x, y) to $(x + 1, y)$ and to $(x, y + 1)$ are respectively λ_x, μ_x $(\lambda_x + \mu_x = 1)$, which in general depend on x. For example, in Bailey's problem:

$$\lambda_x = \frac{n-x}{n-x+\rho}, \quad \mu_x = \frac{\rho}{n-x+\rho}.$$

The motion starts at $(0, 0)$, and stops as soon as any point of the barriers,

$$x = n, \quad y = x + a,$$

is reached. This corresponds to the end of the epidemic, and we are interested in the probability P_w that the particle stops on $(w, a + w)$ $(w = 0, 1, ..., n - 1)$. In the general case we define

$$P_n \equiv 1 - \sum_{w=0}^{n-1} P_w.$$

* See pp. 116–22 above of the present issue.

Thus we have a Birth-and-Death process with the parameters λ_x, μ_x, and the novelty resides in the fact that these parameters are position-dependent. It will be noted, however, that they are assumed independent of the y-co-ordinate. We study the problem for quite general parameters of this type.

3. We now introduce some notation for symmetric functions. Denote by $h_l(x, m)$ the homogeneous product-sum of weight l in the $(m+1)$ quantities, μ_x, μ_{x+1}, $\ldots$, μ_{x+m} (cf. MacMahon, 1915). When these quantities are assumed distinct, define

$$h_l^{(j)}(x, m) = \mu_{x+j}^{l+m}/(\mu_{x+j}-\mu_x)\cdots(\mu_{x+j}-\mu_{x+j-1})(\mu_{x+j}-\mu_{x+j+1})\cdots(\mu_{x+j}-\mu_{x+m}).$$

Then we have the formula
$$h_l(x, m) = \sum_{j=0}^{m} h_l^{(j)}(x, m).$$

Now define
$$q_l(x, m) = \lambda_x \lambda_{x+1} \cdots \lambda_{x+m-1} h_l(x, m),$$

and
$$q_l^{(j)}(x, m) = \lambda_x \lambda_{x+1} \cdots \lambda_{x+m-1} h_l^{(j)}(x, m).$$

In this notation, it may be verified that Whittle's set of recurrence relations becomes

$$r_k: \quad \sum_{i=0}^{k} q_i^{(i)}(k-i, i) P_{k-i} = q_{a+k}^{(k)}(0, k) \quad (k = 0, 1, \ldots, n-1).$$

4. As Whittle observes, in these relations we have to assume that the μ's are all distinct. We proceed now to derive the general formula, using a simple probability argument.

When no barriers are present, the probability that the particle attains the point $(x+m, y+l)$ from the point (x, y) is $q_l(x, m)$. Now P_k is the probability of attaining $(k, a+k)$ from $(0, 0)$ *by any path below the barrier*. Therefore

$$P_k = q_{a+k}(0, k) - q_k(0, k) P_0 - q_{k-1}(1, k-1) P_1 \ldots - q_1(k-1, 1) P_{k-1}.$$

We may rewrite this as the set of relations,

$$\rho_k: \quad \sum_{i=0}^{k} q_i(k-i, i) P_{k-i} = q_{a+k}(0, k) \quad (k = 0, 1, \ldots, n-1).$$

This formula is valid quite generally, and Whittle's formula may be obtained from it by use of the following relations, which are easily verified:

$$\rho_k = \sum_{i=0}^{k} q_i^{(0)}(k-i, i) r_{k-i} \quad (k = 0, 1, \ldots, n-1).$$

5. As an example, we consider the special case where the λ's and μ's are constant:

$$\lambda_x \equiv \lambda, \quad \mu_x \equiv \mu.$$

Then
$$q_r(s, t) = \lambda^t \mu^r \binom{r+t}{t}.$$

Thus
$$\sum_{i=0}^{k} (\lambda\mu)^i \binom{2i}{i} P_{k-i} = \lambda^k \mu^{a+k} \binom{a+2k}{k} \quad (k = 0, 1, \ldots, n-1),$$

which has the solution
$$P_k = \lambda^k \mu^{a+k} \frac{a}{a+2k} \binom{a+2k}{k}.$$

It is interesting to note the connexion between the stochastic epidemic problem and results related to the arc sine law in the theory of fluctuations in coin-tossing (cf. Feller, 1950, p. 252). Thus the case of constant λ, μ may be applied to the tossing of a biased coin with probabilities λ, μ of heads or tails. Then P'_k is the probability that the number of heads is greater than or equal to the number of tails plus a for the first time at the $(2k+a)$th trial. The probability

$$P_n = 1 - \sum_{k=0}^{n-1} P_k$$

is interpreted as the probability that in $(2n+a-1)$ trials the number of heads is always less than the number of tails plus a. For the particular case $a = 1$, $\lambda = \mu = \frac{1}{2}$, we have thus

$$P_n = 1 - \sum_{k=0}^{n-1} \frac{1}{2^{2k+1}} \frac{1}{2k+1} \binom{2k+1}{k},$$

and it may be verified that this equals $2^{-2n} \binom{2n}{n}$, in agreement with the formula of Theorem 1 in Feller (1950, p. 252).

REFERENCES

BAILEY, NORMAN T. J. (1953). The total size of a stochastic epidemic. *Biometrika*, **40**, 177.

FELLER, WILLIAM (1950). *An Introduction to Probability Theory and its Applications*. New York: John Wiley and Sons.

MACMAHON, P. A. (1915). *Combinatory Analysis*. Cambridge University Press.

WHITTLE, P. (1955). The outcome of a stochastic epidemic —a note on Bailey's paper. *Biometrika*, **42**, 116–22.

THE DETERMINISTIC MODEL OF A SIMPLE EPIDEMIC FOR MORE THAN ONE COMMUNITY

By S. RUSHTON and A. J. MAUTNER

Imperial College, London

1. Introduction

In general, a deterministic model can be expected to give a satisfactory picture of a developing process such as the spread of an epidemic, or the growth of a population, so long as the numbers of individuals are sufficiently large. On the other hand, if the numbers are small the continuous description is no longer valid; moreover, the effects of chance occurrences become appreciable in any particular instance, so that a probability treatment is necessary. The literature on deterministic and stochastic models for epidemics has been reviewed by Bailey (1952); to the references given there should be added Bailey 1953 a, b). Much of this previous work refers essentially to a single community or to a set of isolated communities. In the present note we consider the deterministic model of a simple epidemic for several related communities. A simple epidemic is one in which it is assumed that infection spreads only by contact between individuals, and that none of the infected individuals is removed from circulation by death, recovery or isolation. We do not suggest that the model we discuss here is a realistic one in the sense that any actual epidemic has followed precisely this over-simplified pattern, but all epidemic models so far discussed have lacked realism in some sense. We do suggest, however, that the model discussed in this note can provide a basis for further work along these particular lines so that more realistic and therefore more complicated models can be treated by starting from our basic solution.

2. The differential equations for m communities

We denote by $y_i(t)$ the number of susceptibles in the ith community at time t and by n_i the total size of the ith community $(i = 1, ..., m)$. Then, on the assumptions (i) that there is homogeneous mixing within each community with α_i as the internal infection rate in the ith community and (ii) that there is also homogeneous mixing between communities with β_{ij} as the infection rate between the ith and jth communities, the differential equations for the y_i's are

$$\frac{dy_i}{dt} = -y_i\{\alpha_i(n_i - y_i) + \sum_{j \neq i} \beta_{ij}(n_j - y_j)\} \quad (i = 1, ..., m). \tag{1}$$

Under the simplifying assumptions that $n_i = n$ and $\alpha_i = \alpha$ (all i), and that $\beta_{ij} = \alpha\gamma$ (all $i \neq j$) and taking αt as the time-scale and continuing to denote it by t, these equations become

$$\frac{dy_i}{dt} = -y_i\{(n - y_i) + \gamma \sum_{j \neq i} (n - y_j)\} \quad (i = 1, ..., m). \tag{2}$$

The parameter γ in these equations is then the ratio of the assumed common 'cross-infection rate' between communities to the common internal infection rate α, which takes account only of internally generated infection. The solutions of the set (2) are completely

determined by m independent initial conditions, e.g the m values $y_i(0)$ $(i = 1, ..., m)$, and it follows that if any set of these initial values are equal then the corresponding functions $y_i(t)$ are identical for all t. In particular, if all the $y_i(0)$ are equal, the set (2) reduce to a single equation

$$dy/dt = -(1 + (m-1)\gamma)(n-y)y$$

for $y(t) = y_1(t) = ... = y_m(t)$. This is just the equation for a simple epidemic in a single community with total infection rate $\alpha(1 + (m-1)\gamma)$.

Another particular case, whose solution is considered below in §§ 4 and 5, occurs when $(m-1)$ of the initial values $y_i(0)$ are equal and the remaining one, say $y_1(0)$, is not equal to the others. For example, this case arises if a single infected person appears initially in only one of the communities (say the first), so that

$$y_1(0) = n-1 \quad \text{and} \quad y_2(0) = y_3(0) = ... = y_m(0) = n.$$

The equations (2) in such a case reduce to the pair of equations

$$\left.\begin{array}{l} dy_1/dt = -y_1(n - y_1 + (m-1)\gamma(n - y_2)). \\ dy_2/dt = -y_2(\gamma(n - y_1) + (1 + (m-2)\gamma)(n - y_2)), \end{array}\right\} \tag{3}$$

since $y_2(t) = y_3(t) = ... = y_m(t)$ for all t.

3. A general solution

If the initial values $y_i(0)$ $(i = 1, ..., m)$ are all distinct we can obtain a solution of the equations (2) as follows. Writing equations (2) in the form

$$\frac{dy_i}{dt} = y_i(y_i + \gamma \sum_{j \neq i} y_j - na) \quad (i = 1, ..., m), \tag{4}$$

where $a = 1 + (m-1)\gamma$, and making the transformations

$$y_i = nY_i, \quad s = nt, \quad Y_i = e^{-as}U_i, \tag{5}$$

we obtain the equations

$$\frac{1}{U_i}\frac{dU_i}{ds} = e^{-as}(U_i + \gamma \sum_{j \neq i} U_j) \quad (i = 1, ..., m). \tag{6}$$

Changing the independent variable to

$$v = (1 - e^{-as})/a, \tag{7}$$

these equations become

$$\frac{1}{U_i}\frac{dU_i}{dv} = \frac{d}{dv}\log U_i = U_i + \gamma \sum_{j \neq i} U_j \quad (i = 1, ..., m). \tag{8}$$

Denoting the matrix of coefficients on the right of these equations by $\mathbf{A}$, the vector of elements U_i by $\mathbf{U}$, and denoting a column vector by $\{\ \}$, this set is

$$\frac{d}{dv}\{\log U_i\} = \mathbf{A}\mathbf{U}. \tag{9}$$

It follows that

$$\frac{d}{dv}\mathbf{A}^{-1}\{\log U_i\} = \frac{d}{dv}\left\{\log \prod_{j=1}^{m} U_j^{a^{ij}}\right\} = \mathbf{U}, \tag{10}$$

where, corresponding to $\mathbf{A} = [a_{ij}]$, we denote $\mathbf{A}^{-1} = [a^{ij}]$.

Now, putting

$$X_i = \prod_{j=1}^{m} U_j^{a_{ij}}, \tag{11}$$

so that

$$U_i = \prod_{j=1}^{m} X_j^{a_{ij}}, \tag{12}$$

the equations (10) become

$$\frac{d}{dv} \log X_i = \frac{1}{X_i} \frac{dX_i}{dv} = \prod_{j=1}^{m} X_j^{a_{ij}} = \left(\prod_{j=1}^{m} X_j \right)^{\gamma} X_i^{1-\gamma} \tag{13}$$

for $i = 1, ..., m$. We then have

$$X_i^{\gamma-1} \frac{dX_i}{dv} = \left(\prod_{j=1}^{m} X_j \right)^{\gamma} = F(v), \quad \text{say}, \tag{14}$$

for all i. Integrating these equations we obtain

$$X_i = (\alpha_i - G)^{-1/(1-\gamma)} \quad (i = 1, ..., m), \tag{15}$$

where

$$G(v) = (1 - \gamma) \int_0^v F(v)\, dv, \tag{16}$$

and

$$\alpha_i = (X_i(0))^{\gamma-1}. \tag{17}$$

From (12) and the definition of $F(v)$ in (14) we have

$$U_i = \left(\prod_{j=1}^{m} X_j \right)^{\gamma} X_i^{1-\gamma} \tag{18}$$

$$= F(v)/(\alpha_i - G) \quad (i = 1, ..., m) \tag{19}$$

from (15).

Since, from (14), (15) and (16),

$$F(v) = \left\{ \prod_{i=1}^{m} (\alpha_i - G) \right\}^{-\gamma/(1-\gamma)} = \frac{1}{1-\gamma} \frac{dG}{dv}, \tag{20}$$

we may obtain an integral expression for v as

$$v = \frac{1}{1-\gamma} \int_0^G \left\{ \prod_{i=1}^{m} (\alpha_i - G) \right\}^{\gamma/(1-\gamma)} dG. \tag{21}$$

By means of (19) and (21), using the expression for $F(v)$ in terms of G in (20), the solutions for U_i and the time parameter v are given parametrically in terms of G.

4. THE SPECIAL CASE OF EQUATIONS (3)

In the special case mentioned at the end of § 2 in which $y_2(0) = \ldots = y_m(0)$ and $y_1(0) \neq y_2(0)$, the general equations (2) reduce to the pair of equations (3).

We see from (5) and (18) that these conditions imply that

$$U_2(0) = U_3(0) = \ldots = U_m(0), \quad U_1(0) \neq U_2(0)$$

and

$$X_2(0) = X_3(0) = \ldots = X_m(0), \quad X_1(0) \neq X_2(0).$$

Defining

$$(X_1(0))^{\gamma-1} = \alpha, \tag{22}$$

$$(X_i(0))^{\gamma-1} = \beta \quad \text{for} \quad i = 2, 3, \ldots, m, \tag{23}$$

it follows from (16) that

$$U_1 = F(v)/(\alpha - G) \tag{24}$$

and

$$U_i = F(v)/(\beta - G) \quad (i = 2, ..., m). \tag{25}$$

Now put
$$Z = U_1/U_2, \quad W = U_2 - U_1, \tag{26}$$

so that
$$U_2 = W/(1-Z), \quad U_1 = WZ/(1-Z), \tag{27}$$

and note that if $y_2(0) > y_1(0)$, which we may take to be the case, then $y_2(t) > y_1(t)$ for all t, and it follows that $U_2 > U_1$, so that $W \geqslant 0$ and $0 \leqslant Z \leqslant 1$ for all t. Also we must have $\alpha > \beta$.

Expressing W in terms of Z by means of the above equations and the expression for $F(v)$ in (20), we obtain
$$W = A\{(1-Z)^\mu/Z^{\mu-1}\}^\lambda, \tag{28}$$

where $\lambda = 1/(1-\gamma)$, $\mu = 2 + (m-2)\gamma$ and A is an arbitrary constant $= (\alpha - \beta)^{\lambda(1-\mu)}$.

From (20) we have
$$\frac{dv}{dG} = \frac{\lambda}{F(v)}.$$

It is easily seen that this is equivalent to
$$\frac{dv}{dZ} = -\frac{\lambda}{ZW}, \tag{29}$$

so that, corresponding to (21), we may express v in terms of Z as
$$v = B - \frac{\lambda}{A}\int_0^Z \{(1-Z)^\mu/Z^{\mu-1}\}^{-\lambda}\, dZ/Z, \tag{30}$$

where B is a second arbitrary constant. The integral in (30) is an Incomplete Beta Function which can be expressed in terms of the hypergeometric function. In practice, however, it is preferable to evaluate the integral in (30) directly.

5. Practical form of the solution

Considering the practical form of the solution of § 4, we note that if we choose a new time-parameter
$$T = (1 + (m-1)\gamma)v = (1 - e^{-(1+(m-1)\gamma)s}), \tag{31}$$

where we recall that we have put $s = nt$, then $0 \leqslant T \leqslant 1$ and the whole history of the epidemic is shown on a standard time scale extending over a unit interval. Equations (28) and (30) constitute the solution to our problem, the solution being expressed parametrically in terms of $Z = U_1/U_2 = Y_1/Y_2 = y_1/y_2$.

Defining
$$g(Z) = \{(1-Z)^\mu/Z^{\mu-1}\}, \tag{32}$$

then, from (28), (30) and (31),
$$W = Ag(Z), \tag{33}$$

and
$$T = C - \lambda(\mu-1)\int_0^Z (1/WZ)\, dZ, \tag{34}$$

$$= C - (\lambda(\mu-1)/A)\int_0^Z dZ/(Zg(Z)). \tag{35}$$

The initial values of W and Z are
$$W_0 = U_2(0) - U_1(0) = Y_2(0) - Y_1(0) = (y_2(0) - y_1(0))/n, \tag{36}$$

and $Z_0 = y_1(0)/y_2(0)$. We notice incidentally that if $y_2(0) = n$, so that initially there are no infected persons in any community other than the first, then $Z_0 = 1 - W_0$.

The arbitrary constants A and C are clearly given by
$$A = W_0/g(Z_0) \tag{37}$$

and
$$C = (\lambda(\mu-1)/A)\int_0^{Z_0} dZ/(Zg(Z)). \tag{38}$$

9

From equations (5) and (31) the proportions Y_1 and Y_2 of susceptibles in the communities at any time are

$$Y_2 = (1-T)\,U_2$$
$$= (1-T)\,W/(1-Z) = A(1-T)\,g(Z)/(1-Z) \tag{39}$$

in virtue of (27), and
$$Y_1 = ZY_2. \tag{40}$$

Table 1. *Numerical procedure:* $\gamma = 0\cdot2$, $m = 2$, $Z_0 = 0\cdot95$

Z	$g(Z)$	$\displaystyle\int_0^Z \frac{dZ}{Zg(Z)}$	T	s	Y_2	Y_1	$-\dfrac{dY_2}{ds}$	$-\dfrac{dY_1}{ds}$	$-\left(\dfrac{dY_1}{ds}+\dfrac{dY_2}{ds}\right)$
0·95	$5\cdot96 \times 10^{-4}$	57·03	0·000	0·00	1·000	0·950	0·010	0·047	0·057
0·94	$9\cdot53 \times 10^{-4}$	42·94	0·252	0·24	0·996	0·937	0·016	0·060	0·077
0·93	$1\cdot42 \times 10^{-3}$	33·73	0·417	0·45	0·992	0·923	0·023	0·073	0·096
0·92	$2\cdot01 \times 10^{-3}$	27·32	0·531	0·63	0·987	0·908	0·031	0·086	0·117
0·91	$2\cdot73 \times 10^{-3}$	22·65	0·615	0·80	0·981	0·893	0·039	0·099	0·138
0·90	$3\cdot61 \times 10^{-3}$	19·13	0·678	0·94	0·975	0·877	0·049	0·112	0·161
0·89	$4\cdot64 \times 10^{-3}$	16·40	0·727	1·08	0·968	0·862	0·057	0·123	0·181
0·88	$5\cdot85 \times 10^{-3}$	14·23	0·765	1·21	0·959	0·844	0·069	0·138	0·207
0·87	$7\cdot25 \times 10^{-3}$	12·47	0·797	1·33	0·951	0·827	0·080	0·151	0·231
0·86	$8\cdot85 \times 10^{-3}$	11·03	0·823	1·44	0·941	0·809	0·092	0·164	0·256
0·85	$1\cdot07 \times 10^{-2}$	9·83	0·844	1·55	0·930	0·791	0·104	0·177	0·280
0·84	$1\cdot27 \times 10^{-2}$	8·81	0·862	1·66	0·915	0·769	0·120	0·191	0·310
0·83	$1\cdot50 \times 10^{-2}$	7·95	0·878	1·75	0·907	0·753	0·129	0·200	0·330
0·82	$1\cdot76 \times 10^{-2}$	7·20	0·891	1·85	0·893	0·733	0·143	0·212	0·355
0·81	$2\cdot05 \times 10^{-2}$	6·56	0·903	1·94	0·880	0·713	0·157	0·222	0·379
0·80	$2\cdot36 \times 10^{-2}$	5·99	0·913	2·03	0·865	0·692	0·170	0·232	0·402
0·75	$4\cdot48 \times 10^{-2}$	3·99	0·949	2·47	0·774	0·580	0·240	0·270	0·510
0·70	$7\cdot70 \times 10^{-2}$	2·81	0·970	2·91	0·654	0·458	0·297	0·280	0·577
0·65	$1\cdot24 \times 10^{-1}$	2·05	0·983	3·41	0·500	0·325	0·318	0·252	0·569
0·60	$1\cdot92 \times 10^{-1}$	1·53	0·992	4·08	0·301	0·181	0·260	0·173	0·433
0·55	$2\cdot87 \times 10^{-1}$	1·16	0·999	5·89	0·046	0·025	0·052	0·029	0·081

In terms of Y_1 and Y_2 and the time parameter s $(0 \leqslant s \leqslant \infty)$, where

$$s = \{\log_e (1-T)\}/(1+(m-1)\,\gamma), \tag{41}$$

the epidemic rate $-dY_1/ds$ in the first community and $-dY_2/ds$ in the other communities are given by

$$\left.\begin{aligned}
dY_1/ds &= -Y_1(Y_1 + (m-1)\,\gamma Y_2 - (1+(m-1)\,\gamma)),\\
dY_2/ds &= -Y_2(\gamma Y_1 + (1+(m-2)\,\gamma)\,Y_2 - (1+(m-1)\,\gamma)),
\end{aligned}\right\} \tag{42}$$

and the total epidemic rate is $-(dY_1/ds + (m-1)\,dY_2/ds)$.

For communities of total size n the actual epidemic rates are just $-dy_i/dt = -n^2\,dY_i/ds$.

The numerical procedure in applying the solution is illustrated in Table 1 for the simple case of $m = 2$ communities with $\gamma = 0\cdot2$ and initially 5 % of infected individuals in the first community and none in the second. In this case $Y_1(0) = 0\cdot95$, $Y_2(0) = 1$, $W_0 = 0\cdot05$, $Z_0 = 0\cdot95$.

$\lambda = \frac{5}{4}$ and $\mu = 2$. The arbitrary constants are $A = 83 \cdot 8712$ and $C = 1 \cdot 0199$. The principal part of the numerical work is the computing of $g(Z)$ and the integral $\int_0^Z dZ \, (Zg(Z))$. In this example, the epidemic is effectively completed by the time that Z has decreased in value to $0 \cdot 55$, when $T = 0 \cdot 999$, and the values of $g(Z)$ and $\int_0^Z dZ \, (Zg(Z))$ are therefore only shown for $0 \cdot 95 \geqslant Z \geqslant 0 \cdot 55$. If the epidemic rates $-dY_1 \, ds$ and $-dY_2 \, ds$ are plotted the curve for $-dY_2/ds$ is at first below that for $-dY_1/ds$ as we should expect, crosses the latter at about its maximum, rises to a slightly higher maximum and remains above the curve for $-dY_1/ds$ until the completion of the epidemic. The curve for the total epidemic rate falls somewhat more slowly from its maximum than the rate at which it rises. We have examined a number
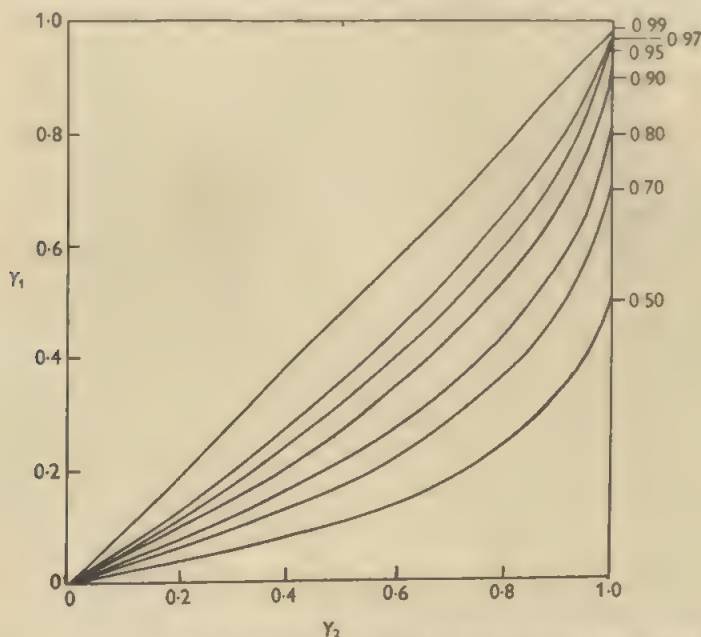
Fig. 1. (Y_1, Y_2) curves for $m = 2$ and $\gamma = 0 \cdot 2$ and different initial values: $Y_1(0) = 0 \cdot 99, 0 \cdot 97, 0 \cdot 95, 0 \cdot 90, 0 \cdot 80, 0 \cdot 70$ and $0 \cdot 50$ and $Y_2(0) = 1$ in all cases.

of cases, for different numbers m of communities and different values of γ, and this last feature appears to be generally true and is particularly noticeable for small m (e.g. $m = 2$) and reasonably large values of the initial amount of infection in one community (e.g. above 10%) before infection in that community begins to affect susceptibles in the others. This calls to mind the remark of Ross (1916): 'It is obvious from the mere examination of many curves of epidemics that they are often remarkably symmetrical bell-shaped curves which, however, frequently tend to fall somewhat more slowly than they rise....'

A method of illustrating the solutions of the epidemic equations is shown in Fig. 1, where for different initial conditions (Y_1, Y_2) curves are plotted for the case $m = 2$ and $\gamma = 0 \cdot 2$. The curves shown are for $Y_1(0) = 0 \cdot 99, 0 \cdot 97, 0 \cdot 95, 0 \cdot 90, 0 \cdot 80, 0 \cdot 70$ and $0 \cdot 50$, and $Y_2(0) = 1$ in all cases. Contours for the time parameter s, giving the times that points on the (Y_1, Y_2) curves are reached from the start of the epidemic, and contours showing the value of the total epidemic rate as the epidemic proceeds can also be put on this same graph. To avoid

confusion they have not been included in Fig. 1, but the latter are in fact conics. In this particular example, we see from equations (42) with $m = 2$, that the contours

$$-(dY_1/ds + dY_2/ds) = k$$

are concentric ellipses

$$Y_1^2 + 2\gamma Y_1 Y_2 + Y_2^2 - (1+\gamma)(Y_1+Y_2) = k,$$

having their common centre at the point $Y_1 = Y_2 = \frac{1}{2}$, their minor axes along the line $Y_1 = Y_2$ of lengths $\sqrt{\{(1+\gamma+2k)/(2(1+\gamma))\}}$ and their major axes (perpendicular to $Y_1 = Y_2$) of lengths equal to $\sqrt{\{(1+\gamma)(1-\gamma)\}}$ times the corresponding minor axes.

REFERENCES

BAILEY, N. T. J. (1952). *Appl. Statist.* **1**, 149.
BAILEY, N. T. J. (1953a). *Biometrika*, **40**, 177.
BAILEY, N. T. J. (1953b). *Biometrika*, **40**, 279.
ROSS, R. (1916). *Proc. Roy. Soc.* A, **92**, 204.

EXACT TESTS FOR SERIAL CORRELATION

By E. J. HANNAN

Australian National University, Canberra, A.C.T.

1. INTRODUCTION

In testing for first-order serial correlation in a stationary time series it is customary to use some form of the first sample serial correlation coefficient. Though this statistic will, presumably, be a most efficient estimator of the serial correlation ρ it has two disadvantages: (*a*) its exact distribution is very complicated so that approximations have to be used; (*b*) it is a biased estimator and the analytical form of the bias is not known.

Ogawara (1951) has obtained an exact test for serial correlation by considering the conditional distribution of the x_{2i} for fixed values of the x_{2i-1}. His statistic $\hat{b}_1$ is an unbiased estimator of $2\rho(1+\rho^2)^{-1}$, where ρ is the parameter of a simple Markoff process. At first sight it seems obvious that the resulting estimator of ρ will be inefficient. However, in this paper it will be shown that its efficiency tends to 1 as ρ tends to 0. It follows that Ogawara's statistic leads to an exact test of the hypothesis which is asymptotically fully efficient (this criterion is defined and justified below). Computationally the test is as simple as that based on the first serial correlation coefficient.

Another statistic, $\hat{b}_2$, is available which is also an unbiased estimator of $2\rho(1+\rho^2)^{-1}$. This statistic is the coefficient of regression of the x_{2i-1} on the neighbouring x_{2i}. The efficiency of the estimator of ρ based on the statistic $\frac{1}{2}(\hat{b}_1 + \hat{b}_2)$ is here shown to be $(1-\rho^2)$. While no exact distribution theory is available for this statistic it could be useful in circumstances where an estimate of a common serial correlation coefficient is required from a number of otherwise differing simple Markoff processes, since it is also an unbiased estimator of $2\rho(1+\rho^2)^{-1}$.

Ogawara also gave the extension of his results to the case of a multiple Markoff process of order h. In general, however, it appears that the efficiency of his estimates of the coefficients of the autoregressive equation will not rise above $2/(h+1)$, so that only in the first-order case are they ever fully efficient.

The problem of testing for serial correlation in the residuals from a regression equation has been considered by Moran (1950) and Durbin & Watson (1950, 1951). Moran obtained asymptotic formulae for the mean and variance of the first circular serial correlation coefficient computed from the residuals from a least-squares regression on a single independent variate. The exact distribution of the test statistic used by Durbin & Watson is of a complicated form and is only known for regression vectors satisfying certain conditions. However, upper and lower bounds to the significance points, valid for any regression vectors, have been tabulated. In cases where the test statistic falls between the appropriate bounds an approximate test was suggested, based on the use of a beta distribution with the same mean and variance as the test statistic.

In this paper it will be shown that Ogawara's method can be extended to give an exact test, for the independence of the residuals from the regression equation, based on the z distribution. Again, a statistic is obtained which is an unbiased estimator of $2\rho(1+\rho^2)^{-1}$ in the case where the residuals follow a first-order Markoff scheme. The estimator is, as ρ

tends to zero, asymptotically as efficient an estimator of ρ as the first serial correlation coefficient of the residuals and Durbin & Watson's statistic. It again follows that the statistic leads to a test which is asymptotically fully efficient.

At the same time an exact test against given values of the regression coefficients is obtained in the case in which the residuals follow a simple Markoff process.

One disadvantage of these tests results from the necessity of computing a separate least-squares regression involving $2k+1$ regressors, where k is the number of 'independent' variates in the original regression equation. The estimates of the coefficients in the original equation, obtained from this regression, will, under some circumstances, have an asymptotically smaller variance than the estimates obtained by straightforward least squares. However, over a wide range of these conditions the estimates of the regression coefficients obtained from the first differences of the observations will have an even smaller asymptotic variance, and over this range at least the variances of these last estimates will almost certainly be smaller for small samples.

Finally, it should be emphasized that, though certain exact tests have been obtained, the powers of these tests have been judged from their asymptotic properties. For really small samples these tests may be far from optimal.

2. The criterion of asymptotic relative efficiency of tests

If t_1 and t_2 are statistics, computed from a sample of size n, their asymptotic efficiency as tests of an hypothesis specified by a parameter value θ_0 is defined as

$$E(t_1, t_2 \mid \theta_0) = \lim_{n \to \infty} \frac{\left|\left[\frac{\partial}{\partial \theta} \mathscr{E}(t_1)\right]_{\theta=\theta_0}\right|^2}{V(t_1 \mid \theta = \theta_0)} \frac{V(t_2 \mid \theta = \theta_0)}{\left|\left[\frac{\partial}{\partial \theta} \mathscr{E}(t_2)\right]_{\theta=\theta_0}\right|^2}.$$

Here $\mathscr{E}(t_j)$ and $V(t_j)$ are respectively the expected value and variance of t_j.

The justification for this criterion rests upon a theorem due to Pitman (see Stuart, 1954). Pitman has shown that, under certain regularity conditions, for t_1 and t_2 with limiting normal distributions and variances of order n^{-1}, the ratio of the sample sizes required for t_1 and t_2 to have the same power against alternative values of θ which differ from θ_0 by quantities of order $n^{-\frac{1}{2}}$ is in the limit given by $E(t_1, t_2 \mid \theta_0)$.

3. The test for serial correlation in a stationary autoregressive process

Ogawara considered the stochastic process $(x_t - m) = \rho(x_{t-1} - m) + \epsilon_t$, where ϵ_t is normally distributed with 0 mean and variance $\sigma^2(1 - \rho^2)$, and m is the mean of the process. Here $|\rho| < 1$.

He showed that in the conditional distribution of the x_{2t} ($t = 1, \ldots, n$) for fixed values of x_{2t-1} ($t = 1, \ldots, n+1$), the parameter $b = 2\rho(1+\rho^2)^{-1}$ appears as the regression coefficient of the x_{2t} on the fixed variates $\frac{1}{2}(x_{2t-1} + x_{2t+1}) = x_t'$. Ogawara pointed out that the statistic

$$\hat{b}_1 = \frac{\sum_{1}^{n}(x_t' - \bar{x}')(x_{2t} - \bar{x}_2)}{\sum(x_t' - x')^2}$$

is then the maximum-likelihood estimator of b, and is unbiased, and that

$$F = \frac{(\hat{b}_1 - b)^2 \sum_1^n (x_t' - \bar{x}')^2}{\sum_1^n (x_{2t} - \hat{a} - \hat{b}_1 x_t')^2} (n-2),$$

which has the F distribution with 1 and $n-2$ degrees of freedom, can be used as an exact test for any assigned value of b. Here $\hat{a} = \bar{x}_2 - \hat{b}_1 x'$.

For the test of significance of the hypothesis $\rho = 0$ the statistic F reduces to

$$(n-2) r^2 (1 - r^2)^{-1},$$

where r is the coefficient of correlation between x_{2t} and $\frac{1}{2}(x_{2t-1} + x_{2t+1})$. This test is, therefore, computationally as simple as that based on the first serial correlation coefficient.

The variance of $\hat{b}_1$, in the conditional distribution of the x_{2t} is

$$\frac{\sigma^2(1-\rho^2)}{1+\rho^2} \frac{1}{\sum_1^n (x_t' - \bar{x}')^2},$$

and since this converges in probability as n increases to

$$\frac{\sigma^2(1-\rho^2) 4}{(1+\rho^2) 2n\sigma^2(1+\rho^2)} = \frac{2}{n} \frac{(1-\rho^2)}{(1+\rho^2)^2},$$

the conditional variance of $\hat{b}_1$ also converges in probability to[†]

$$\frac{2}{n} \frac{1-\rho^2}{(1+\rho^2)^2}.$$

Given $\hat{b}_1$ we can estimate ρ by

$$\hat{\rho}_1 = \frac{1 - \sqrt{(1-\hat{b}_1^2)}}{\hat{b}_1}.$$

This is not, of course, unbiased. Its variance will be, in the limit,

$$\frac{2}{n} \frac{1-\rho^2}{(1+\rho^2)^2} \frac{(1+\rho^2)^4}{4(1-\rho^2)^2} = \frac{1}{2n} \frac{(1+\rho^2)^2}{(1-\rho^2)}.$$

The variance of r_1, the first sample serial correlation coefficient, will tend to $\frac{1}{2}(1-\rho^2)/n$ (Bartlett, 1946), and since it seems that this will be asymptotically most efficient (Wald, 1948) the efficiency of Ogawara's estimate, $\hat{\rho}_1$, will be

$$\left(\frac{1-\rho^2}{1+\rho^2}\right)^2.$$

The efficiency of $\hat{\rho}_1$ for certain values of ρ is shown in Table 1.

Table 1

$|\rho|$	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
Efficiency of $\hat{\rho}_1$	0·96	0·85	0·70	0·52	0·36	0·22	0·12	0·05	0·01

[†] For any observed $\hat{b}_1$ the variance, for fixed x_{2t}, will be $\sigma^2(1-\rho^2)\left\{(1+\rho^2)\sum_1^n (x_t' - \bar{x}')^2\right\}^{-1}$ of course.

Another unbiased estimator of b is available in the statistic

$$\hat{b}_2 = \frac{\sum_1^{n-1} (x_i'' - \bar{x}'') x_{2i+1}}{\sum_1^{n-1} (x_i'' - \bar{x}'')^2},$$

where $x_i'' = \frac{1}{2}(x_{2i} + x_{2i+2})$.

The statistic $\hat{b} = \frac{1}{2}(\hat{b}_1 + \hat{b}_2)$ will also be unbiased and its variance will be

$$\tfrac{1}{4}\operatorname{var}(\hat{b}_1) + \tfrac{1}{4}\operatorname{var}(\hat{b}_2) + \tfrac{1}{2}\operatorname{cov}(\hat{b}_1\hat{b}_2) = \frac{1-\rho^2}{n(1+\rho^2)^2} + \tfrac{1}{2}\operatorname{cov}(\hat{b}_1\hat{b}_2).$$

To the order n^{-1} the sample means can be neglected and the covariance of $\hat{b}_1$ and $\hat{b}_2$ will be (to this order):

$$\frac{1}{(1+\rho^2)^2}\operatorname{cov}(a_1 a_2) + \frac{\rho^2}{(1+\rho^2)^4}\operatorname{cov}(c_1 c_2) - \frac{\rho}{(1+\rho^2)^3}\{\operatorname{cov}(a_1 c_2) + \operatorname{cov}(c_1 a_2)\},$$

where

$$a_1 = \frac{1}{\sigma^2 n}\sum_1^n x_{2i}(x_{2i-1} + x_{2i+1}), \quad a_2 = \frac{1}{\sigma^2 n}\sum_1^{n-1} x_{2i+1}(x_{2i} + x_{2i+2}),$$

$$c_1 = \frac{1}{\sigma^2 n}\sum_1^n (x_{2i-1} + x_{2i+1})^2, \quad c_2 = \frac{1}{\sigma^2 n}\sum_1^{n-1}(x_{2i} + x_{2i+2})^2.$$

By straightforward algebra we obtain, to the order n^{-1},

$$\operatorname{cov}(\hat{b}_1\hat{b}_2) = \frac{1}{n}\left\{ \frac{1}{(1+\rho^2)^2}\left[\frac{2 + 10\rho^2 + 6\rho^4 - 2\rho^6}{1-\rho^4}\right] + \frac{\rho^2}{(1+\rho^2)^4}\left[\frac{40\rho^2 + 40\rho^4 - 8\rho^6 - 8\rho^8}{1-\rho^4}\right]\right.$$
$$\left. - \frac{2\rho}{(1+\rho^2)^3}\left[\frac{12\rho + 20\rho^3 + 4\rho^5 - 4\rho^7}{1-\rho^4}\right]\right\}.$$

The asymptotic variance of $\hat{b}$, therefore, is

$$\operatorname{var}(\hat{b}_1)\left[\frac{1-\rho^2}{(1+\rho^2)^2}\right].$$

Hence, the variance of the estimator $\hat{\rho} = [1 - \sqrt{(1-\hat{b}^2)}]\hat{b}^{-1}$ tends to $(2n)^{-1}$, so that its efficiency is $(1-\rho^2)$.

Table 2 shows the efficiency of this statistic for certain values of ρ.

Table 2

$|\rho|$	0·1	0·2	0·3	0·4	0·5	0·6	0·7	0·8	0·9
Efficiency of $\hat{\rho}$	0·99	0·96	0·91	0·84	0·75	0·64	0·51	0·36	0·19

Though no exact distribution theory is available for this statistic, yet, in spite of its inefficiency, it could prove useful in circumstances where an estimate of a common serial correlation coefficient is required from a number of otherwise differing simple Markoff processes.[†] The lack of bias in $\frac{1}{2}(\hat{b}_1 + \hat{b}_2)$ would enable the several estimates to be combined

[†] An example (due to E. J. Williams) has arisen in connexion with the serial correlation of measurements at points along a log. The averaging of estimates from different logs, the only method of substantially increasing the number of observations, would not result in a consistent estimator if the estimates were obtained from the first sample serial correlation coefficient.

and an estimator of ρ to be obtained which would be consistent. The loss of efficiency, for moderate ρ, would not be large.

The statistics $\hat{b}_1$, $\hat{b}_2$ and $b = \frac{1}{2}(\hat{b}_1 + \hat{b}_2)$ have one major disadvantage however. The use of the formula

$$\hat{\rho}_j = [1 - \sqrt{(1 - \hat{b}_j^2)}]/\hat{b}_j$$

will lead to a consistent estimator of the first serial correlation coefficient only if the underlying process is truly a simple Markoff process.

The fact that the efficiency of $\hat{\rho}_j$ tends to 1 as ρ tends to 0 is important, however, for this shows that the test of significance based on $\hat{b}_j$ ($j = 1, 2$) will be asymptotically fully efficient. Though nothing certain can be said, from this result, about the power of the test in small samples it seems likely that it will be nearly as powerful as that based on r_1 against the range of alternatives represented by simple Markoff processes. The test based on $\hat{b}_j$ has the advantage of being exact.

Ogawara extended his results to a multiple Markoff process of order h. In this case the regression of the variate $x_{t(h+1)}$ on the variates $x'_{pt} = \frac{1}{2}(x_{t(h+1)-p} + x_{t(h+1)-p})$ for $p = 1, \ldots, h$ is considered. In the first-order case the basic test statistic is the correlation coefficient between x_{2t} and $\frac{1}{2}(x_{2t-1} + x_{2t+1})$, and the numerator of this statistic (apart from the factor $\frac{1}{2}$) differs only in the corrections for the means from the numerator of the first sample serial correlation coefficient. The corresponding test statistics in the general case are the total and partial correlation coefficients between $x_{t(h+1)}$ and the x'_{pt}. Now, however, certain cross-products are omitted as compared with the total and partial sample serial correlation coefficients so that the power of the tests is reduced. It seems that the efficiency of the estimates of the coefficients of the autoregressive equation (from which the asymptotic power of the tests can be gauged) will be at a maximum when all of the coefficients are zero. In this case the efficiency is $2(h+1)^{-1}$. In general, it will be lower. For example, the efficiency of the estimate of a_2, in the process $x_t + a_1 x_{t-1} + a_2 x_{t-2} = \epsilon_t$ in case $a_2 = 0$, is $\frac{2}{3}(1 + a_1^2)^{-1}$. It appears, therefore, that although exact tests for partial serial correlation are obtained by Ogawara's approach these tests will be less powerful than those provided by the usual serial coefficients.

4. THE TEST FOR SERIAL CORRELATION IN THE RESIDUALS FROM A REGRESSION EQUATION

Consider the regression $\quad y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + \epsilon_t,$

where $\qquad$ (1) $\epsilon_t = \rho \epsilon_{t-1} + \eta_t$ and η_t is $N(0, \sigma(1 - \rho^2)^{\frac{1}{2}})$; $|\rho| < 1$

$\qquad$ (2) The ϵ_t are independent of the x_{jt}.

Then the conditional distribution of the y_{2t} for fixed $x_{j,t}$ and fixed ϵ_{2t+1} ($t = 0, \ldots, n$) is

$$(2\pi\sigma_0^2)^{-\frac{1}{2}n} \exp{-\frac{1}{2\sigma_0^2} \sum_{t=1}^{n} \left(y_{2t} - \gamma_0 - \sum_{j=1}^{2k+1} \gamma_j z_{jt} \right)^2 \prod_1^n dy_{2t}}.$$

Here $\qquad$ $\sigma_0^2 = \dfrac{\sigma^2(1 - \rho^2)}{1 + \rho^2}, \quad \gamma_0 = \alpha(1 - \gamma_{k+1}),$

$$\gamma_j = \beta_j \quad \text{and} \quad z_{j,t} = x_{j,2t} \quad (j = 1, \ldots, k),$$

$$\gamma_{k+1} = \frac{2\rho}{1 + \rho^2} \quad \text{and} \quad z_{k+1,t} = \frac{1}{2}(y_{2t-1} + y_{2t+1}),$$

$$\gamma_j = -\gamma_{k+1}\beta_j \quad \text{and} \quad z_{j,t} = \frac{1}{2}(x_{j,2t-1} + x_{j,2t+1}) \quad (j = k+2, \ldots, 2k+1).$$

If the information involved in the knowledge that $\gamma_{j+k+1} = -\gamma_{k+1}\gamma_j$ $(j = 1, ..., k)$ is discarded and the coefficients γ_j are estimated by the straightforward least-squares procedure, these estimates will have the usual properties of least-squares estimates in the classic case. They will not be maximum-likelihood estimates. However, the computation of the maximum-likelihood estimates will involve the solution of a system of second-order equations and no exact tests will be available.

Representing the least-squares estimates of the γ_j by $\hat{\gamma}_j$, the hypothesis $\rho = 0$ can be tested by the use of the statistic

$$F_{1,n-2k-2} = (n-2k-2)\frac{\hat{\gamma}_{k+1}^2}{\sigma^{*2}}\frac{|L|}{|L_{k+1,k+1}|}.$$

which has the F distribution with 1 and $n-2k-2$ degrees of freedom. Here σ^{*2} is the variance of the residuals from the least-squares regression of y_{2t} on the z_{jt} while L is the covariance matrix of the z_{jt}: $|L_{k+1,k+1}|$ is the cofactor of the element in the indicated row and column. The test is, of course, equivalent to a test of the partial correlation of y_{2t} and $\frac{1}{2}(y_{2t-1} + y_{2t+1})$ with the effects of the z_{jt} $(j \neq k+1)$ removed.

Tests of significance and confidence limits for the parameters β_j can be obtained from statistics such as

$$F_{1,n-2k-2} = (n-2k-2)\left\{(\hat{\gamma}_j - \beta_j)^2 \frac{|L|}{|L_{jj}|}\right\}\sigma^{*-2} \quad (j = 1, ..., k),$$

which have the F distribution with 1 and $(n-2k-2)$ degrees of freedom. Similarly, the multiple correlation of y and the x_j may be tested by the statistic

$$F_{k,n-2k-2} = \frac{(n-2k-2)}{k}\left\{\sum_{ij=1}^{k}\hat{\gamma}_i\hat{\gamma}_j\frac{|\Lambda_{ij}|}{|\Lambda|}\right\}\sigma^{*-2}.$$

Here Λ is the matrix formed from the first k rows and columns of L^{-1}.

These last tests are exact for any value of ρ, though the power of the tests will depend upon that parameter.

In order to examine the limiting variances and covariances of the estimates of the γ_j it is necessary to obtain the stochastic limit of the matrix L.

If the vector $\{z_{jt}\}$ is written $Z_t = \{Z_{1t}, z_{k+1,t}, Z_{2t}\}$ and the corresponding vector of sample means $\bar{Z} = \{\bar{Z}_1, \bar{z}_{k+1}, \bar{Z}_2\}$, then the covariance matrix L is

$$L = \frac{1}{n}\sum_{t=1}^{n}\begin{bmatrix} (Z_{1t}-\bar{Z}_1)Z'_{1t} & (Z_{1t}-\bar{Z}_1)z_{k+1,t} & (Z_{1t}-\bar{Z}_1)Z'_{2t} \\ (z_{k+1,t}-\bar{z}_{k+1})Z'_{1t} & (z_{k+1,t}-\bar{z}_{k+1})z_{k+1,t} & (z_{k+1,t}-\bar{z}_{k+1})Z'_{2t} \\ (Z_{2t}-\bar{Z}_2)Z'_{1t} & (Z_{2t}-\bar{Z}_2)z_{k+1,t} & (Z_{2t}-\bar{Z}_2)Z'_{2t} \end{bmatrix}.$$

Since $z_{k+1,t} = \alpha + \beta'Z_{2t} + \frac{1}{2}(\epsilon_{2t-1} + \epsilon_{2t+1})$, where β is the vector of regression coefficients, under fairly weak restrictions on the nature of the processes generating the x_{jt}, this matrix will have a limit in probability of the form

$$\begin{bmatrix} X_0 & X_1\beta & X_1 \\ \beta'X'_1 & \frac{1}{2}\sigma^2(1+\rho^2)+\beta'X_2\beta & \beta'X_2 \\ X'_1 & X_2\beta & X_2 \end{bmatrix},$$

where X_0, X_1 and its transpose and X_2 are the stochastic limits of the matrix in the corners of L.

Since the determinantal value of L will be almost everywhere different from 0 (if the joint distribution of the x_j is not singular) the matrix L^{-1} will converge in probability to the inverse of the matrix just written.

Since the covariance matrix of the $\hat{\gamma}_j$ (for $j > 0$) is $\sigma_0^2(nL)^{-1}$, it is readily seen that the variance of $\hat{\gamma}_{k+1}$ converges in probability to

$$\frac{2}{n}\frac{1-\rho^2}{(1+\rho^2)^2}.$$

Hence the variance of the estimator of ρ obtained from this statistic is

$$\frac{1}{2n}\frac{(1+\rho^2)^2}{(1-\rho^2)},$$

as in the case of the estimation of the parameter of a simple Markoff process by Ogawara's method.

The test statistic, d, used by Durbin & Watson, is asymptotically equal to $2(1-r_1)$, where r_1 is the first serial correlation coefficient computed from the actual residuals from a least-squares regression. It can be seen from the formulae (5), (6), (7) and (8) given on p. 164 of their paper (1951) that the variance of this statistic (when $\rho = 0$) will be, asymptotically, $2n^{-1}$ when computed from $2n$ observations. It follows that the estimator of ρ obtained from $\hat{\gamma}_{k+1}$ is as efficient as that gained from the statistic d, when $\rho = 0$. The asymptotic relative efficiency of d and $\hat{\gamma}_{k+1}$ against the hypothesis $\rho = 0$ is therefore unity. It is easy to show that the maximum-likelihood estimator of ρ will have an asymptotic variance $(2n)^{-1}$ (from $2n$ observations) when ρ equals zero. It therefore appears that d and $\hat{\gamma}_{k+1}$ lead to tests of the hypothesis, $\rho = 0$, which are asymptotically fully efficient.

The covariance matrix of the $\hat{\gamma}_j$ $(j = 1, ..., k)$ converges in probability to

$$\left(\frac{\sigma^2}{n}\frac{1-\rho^2}{1+\rho^2}\right)X,$$

where X is the matrix in the top left-hand corner of the inverse of

$$\begin{bmatrix} X_0 & X_1 \\ X_1' & X_2 \end{bmatrix}.$$

If the x_j are serially independent the covariance matrix of the $\hat{\gamma}_j$ $(j = 1, ..., k)$ becomes in the limit

$$\frac{\sigma^2}{n}\frac{1-\rho^2}{1+\rho^2}X_0^{-1}.$$

But these $\hat{\gamma}_j$ are unbiased estimators of the β_j. The covariance matrix of the straightforward least-squares estimates of the β_j is, in this very special case (Wold, 1953, p. 213),

$$\frac{\sigma^2}{2n}X_0^{-1}.$$

The relative efficiency of the estimates of the β_j by the two methods is then $\frac{1}{2}(1+\rho^2)(1-\rho^2)^{-1}$. The estimates obtained from the $\hat{\gamma}_j$ will, in this case, be more efficient than those from least squares when $\rho\sqrt{3} > 1$.

In general, the relative efficiency of the two methods will depend upon the correlogram of the x_{jt} as well as ρ. For example, in the case where there is only one regressor, which follows a simple Markoff process with parameter ρ_1, the variance of $\hat{\gamma}_1$ will be, in the limit,

$$\text{var}\,(\hat{\gamma}_1) = \frac{\sigma^2}{n}\left(\frac{1-\rho^2}{1+\rho^2}\right)\left(\frac{1+\rho_1^2}{1-\rho_1^2}\right)\frac{1}{\sigma_1^2}.$$

where σ_1^2 is the variance of x_1.

The variance of the straightforward least-squares estimate from $2n$ observations will then tend to (Wold, 1953, p. 211)

$$\text{var}\,(\hat{\beta}_1) = \frac{\sigma^2}{2n}\frac{1+\rho\rho_1}{1-\rho\rho_1}\frac{1}{\sigma_1^2}.$$

Table 3. $\text{var}\,(\hat{\beta}_1)/\text{var}\,(\hat{\gamma}_1)$

ρ ρ_1	−0·8	−0·6	−0·4	−0·2	0	0·2	0·4	0·6	0·8
0	2·28	1·06	0·69	0·54	0·50	0·54	0·69	1·06	2·28
0·2	1·52	0·77	0·54	0·46	0·46	0·54	0·75	1·25	2·90
0·4	0·85	0·47	0·36	0·33	0·36	0·42	0·69	1·26	3·20
0·6	0·38	0·24	0·20	0·20	0·24	0·32	0·63	1·06	3·05
0·8	0·11	0·08	0·08	0·09	0·11	0·16	0·29	0·66	2·28

The relative efficiency $\qquad$ $\text{var}\,(\hat{\beta}_1)/\text{var}\,(\hat{\gamma}_1)$

is shown in Table 3. This table will also give the relative efficiencies of the $\hat{\gamma}_j$ $(j = 1, \ldots, k)$ for a k variate regression when each of the x_j follows a simple Markoff process with the same parameter ρ_1.

The limiting variance of $\hat{\gamma}_0$ is $\qquad$ $\dfrac{\sigma^2}{n}\dfrac{1-\rho^2}{1+\rho^2}.$

The corresponding estimate of α is $\hat{\gamma}_0(1-\hat{\gamma}_{k+1})^{-1}$, and this has the asymptotic variance

$$\frac{\sigma^2}{n}\frac{1+\rho}{(1-\rho)^3}\left(1+\rho^2+\frac{2\alpha^2}{\sigma^2}\right).$$

The variance of the straightforward least-squares estimate of α will tend to

$$\frac{\sigma^2}{n}\frac{1+\rho}{1-\rho}.$$

so that the estimate obtained from $\hat{\gamma}_0$ is always relatively inefficient and is very inefficient for values of ρ near 1. In such cases a better estimate than that obtained from $\hat{\gamma}_0$ would be $\hat{\alpha} = \bar{y} - \sum_1^k \hat{\gamma}_j \bar{x}_j$, where the means are obtained from all of the observations.

A common procedure, when positive serial correlation is suspected in the residuals, is to estimate the regression coefficients from the first differences of the y_t and x_{jt} (see Cochrane & Orcutt, 1949). Again the efficiency of this method will depend upon the correlogram of the x_{jt}. When there is only one regressor, which follows a simple Markoff process with

variance σ_1^2 and serial correlation ρ_1, the asymptotic variance of the least-squares estimate of β_1, from the first differences, can be shown to be (when estimated from $2n$ observations)

$$\frac{\sigma^2}{\sigma_1^2}\frac{1}{4n}\frac{(1-\rho)}{(1-\rho_1)}\left\{\frac{3-\rho\rho_1-\rho-\rho_1}{1-\rho\rho_1}\right\}.$$

If $\rho \geqslant 0$ this will be smaller than the variance of the estimator $\hat{\gamma}_1$.

It must be emphasized that the validity of the tests of the regression coefficients, together with the estimates of these coefficients (including the estimate of ρ derived from $\hat{\gamma}_{k+1}$), depends on the specification of the errors in the regression equation as a simple Markoff process being correct. The test of significance of ρ is, of course, always an exact test.

Example. Durbin & Watson (1951) use economic data due to A. R. Prest (1949) to demonstrate their method. The original data, given in their paper, cover the period 1870–1938, and show the logarithm of the consumption of spirits per head in the United Kingdom (y_t), the logarithm of real income per head (x_{1t}) and the logarithm of the relative price of spirits (x_{2t}). The estimates of the coefficients from the least-squares regression of y on x_1 and x_2 are

$$\hat{\beta}_1 = -0\cdot120, \quad \hat{\beta}_2 = -1\cdot228.$$

Their statistic $d = 0\cdot2488$ and is far below the lower bound to the significance point for d at the 1 % level, indicating a highly significant positive serial correlation.

We will apply the method presented in this paper to this example.

The coefficients $\hat{\gamma}_j$ are given by

$$\begin{bmatrix}\hat{\gamma}_1\\\hat{\gamma}_2\\\hat{\gamma}_3\\\hat{\gamma}_4\\\hat{\gamma}_5\end{bmatrix} = \begin{bmatrix}0\cdot307 & 0\cdot490 & -0\cdot651 & 0\cdot295 & 0\cdot495\\0\cdot490 & 1\cdot445 & -1\cdot818 & 0\cdot474 & 1\cdot423\\-0\cdot651 & -1\cdot818 & 2\cdot423 & -0\cdot631 & -1\cdot815\\0\cdot295 & 0\cdot474 & -0\cdot631 & 0\cdot288 & 0\cdot478\\0\cdot495 & 1\cdot423 & -1\cdot815 & 0\cdot478 & 1\cdot417\end{bmatrix}^{-1}\begin{bmatrix}-0\cdot636\\-1\cdot818\\2\cdot401\\-0\cdot621\\-1\cdot804\end{bmatrix} = \begin{bmatrix}0\cdot795\\-0\cdot693\\0\cdot950\\-0\cdot792\\0\cdot630\end{bmatrix}$$

$$\hat{\gamma}_0 = 0\cdot218, \quad \sigma^{*2} = 0\cdot000232.$$

The elements of the 5×5 matrix are $\sum_1^{34}(z_{it}-\bar{z}_i)(z_{jt}-\bar{z}_j)$. For example,

$$0\cdot495 = \sum \tfrac{1}{2}(x_{2,2t-1}+x_{2,2t+1}-\bar{x}_2')(x_{1,2t}-\bar{x}_{1,2}),$$
$$\bar{x}_2' = \tfrac{1}{34}\sum \tfrac{1}{2}(x_{2,2t-1}+x_{2,2t+1}),$$
$$\bar{x}_{1,2} = \tfrac{1}{34}\sum x_{1,2t}.$$

where

All of the parameters $\hat{\gamma}_j$ are significant at the 1 % point. The value of F corresponding to $\hat{\gamma}_3$ equals 308 in fact. However, the 5 % confidence intervals for γ_1 and γ_2 do not include the straightforward least-squares estimates. They are:

$$0\cdot339 \leqslant \gamma_1 \leqslant 1\cdot251, \quad -0\cdot991 \leqslant \gamma_2 \leqslant -0\cdot405.$$

On prior grounds the value of $-0\cdot120$ for $\hat{\beta}_1$ is unacceptable, for it is very unlikely that the demand for spirits would fall as income rises. In the original work of A. R. Prest (1949) the coefficients were estimated from the first differences of the observations and the resulting regression coefficients were

$$\hat{\beta}_1 = 0\cdot736, \quad \hat{\beta}_2 = -0\cdot861,$$

both of which lie within the 5 % confidence intervals for γ_1 and γ_2.

The variates x_1 and x_2 have high positive serial correlation coefficients of about the same value. Together with the high positive serial correlation of the residuals this suggests that the extension of Ogawara's method here presented will give estimators of β_1 and β_2 which will be at least as efficient as those from the straightforward least-squares procedure. This is to some extent borne out by the results.

I wish to thank Prof. P. A. P. Moran for suggesting this subject and its developments to me and for his advice in the research done and in the preparation of this paper. I should also like to thank the referee for a number of suggestions and, in particular, for pointing out the importance of the first difference transformation.

REFERENCES

BARTLETT, M. S. (1946). On the theoretical specification and sampling properties of auto-correlated time series. *J.R. Statist. Soc. Suppl.* **8**, 27.

COCHRANE, D. & ORCUTT, G. H. (1949). Application of least squares regression to relationships containing auto-correlated error terms. *J. Amer. Statist. Ass.* **44**, 32.

DURBIN, J. & WATSON, G. S. (1950, 1951). Testing for serial correlation in least squares regression. I. *Biometrika*, **37**, 409; II. *Biometrika*, **38**, 159.

MORAN, P. A. P. (1950). A test for serial independence of residuals. *Biometrika*, **37**, 178.

OGAWARA, M. (1951). A note on the test of serial correlation coefficients. *Ann. Math. Statist.* **22**, 115.

PREST, A. R. (1949). Some experiments in demand analysis. *Rev. Econ. Statist.* **31**, 33.

STUART, A. (1954). Asymptotic relative efficiencies of distribution-free tests of randomness against normal alternatives. *J. Amer. Statist. Ass.* **49**, 147.

WALD, A. (1948). Asymptotic properties of the maximum likelihood estimate of the unknown parameter of a discrete stochastic process. *Ann. Math. Statist.* **19**, 40.

WOLD, H. (1953). *Demand Analysis*. New York: John Wiley and Sons.

ON THE EFFICIENCY OF PROCEDURES FOR SMOOTHING PERIODOGRAMS FROM TIME SERIES WITH CONTINUOUS SPECTRA

By M. S. BARTLETT AND J. MEDHI

University of Manchester

1. INTRODUCTION

Let X_r $(r = 1, 2, ..., n)$ be a set of consecutive observations from a real stochastic process X_t in discrete time. The process is assumed stationary, at least up to the second order, so that the stochastic average $E\{X_t\}$ is a constant (m). which for convenience is put zero, and the autocovariance function

$$E\{X_t X_{t+s}\} = w_s = \sigma_X^2 \rho_s \qquad (1 \cdot 1)$$

is a function only of the interval s. Then an autocorrelation function ρ_s $(s = 0, \pm 1, \pm 2, ...)$ is a valid one for some stationary process X_t if and only if

$$\rho_s = \int_0^\pi \cos(s\lambda) \, dF(\lambda), \qquad (1 \cdot 2)$$

where $F(\lambda)$, the (integrated) spectrum of the process. has the form of a distribution function defined in $(0, \pi)$ (Wold, 1938). Ignoring the possibility of a third singular component, we have further the relation

$$F(\lambda) = c_1^2 F_1(\lambda) + c_2^2 F_2(\lambda), \qquad (1 \cdot 3)$$

where $F_1(\lambda)$ is a step function, the (integrated) discrete spectrum, and $F_2(\lambda)$ has a non-zero derivative $f_2(\lambda)$, the continuous spectrum. In particular when c_1 is zero (as will now be assumed)

$$\rho_s = \int_0^\pi \cos(s\lambda) f(\lambda) \, d\lambda, \qquad (1 \cdot 4)$$

which has the inversion formula

$$f(\lambda) = \frac{1}{\pi} \sum_{s=-\infty}^{\infty} \rho_s \cos(s\lambda). \qquad (1 \cdot 5)$$

We define

$$g(\lambda) = 2\pi \sigma_X^2 f(\lambda),$$

so that

$$g(\lambda) = 2 \sum_{s=-\infty}^{\infty} w_s \cos(s\lambda). \qquad (1 \cdot 6)$$

The problem of estimating directly the spectral density function $f(\lambda)$ (or $g(\lambda)$), previously considered by Daniell (see the discussion to the paper by Bartlett, 1946), Bartlett (1950) and Grenander (1951), will be discussed further in this paper. Grenander & Rosenblatt (1952, 1954) have recently investigated also the problem of constructing an entire confidence band for the integrated density function $g(\lambda)$; for further suggestions on this problem, which will not be considered here, see Bartlett (1954).

2. ESTIMATES OF THE SPECTRAL DENSITY FUNCTION

The sample autocovariance, which will be defined as

$$C_s = \begin{cases} \dfrac{1}{n} \sum_{r=1}^{n-s} X_r X_{r+s} & (|s| < n), \\ 0 & (|s| \geqslant n), \end{cases} \qquad (2 \cdot 1)$$

is, under wide conditions, a consistent estimate of w_s. If now we substitute C_s for w_s in (1·6), we obtain

$$2 \sum_{s=-n}^{n} C_s \cos(s\lambda) = \frac{2}{n} \sum_{\mu,\nu=1}^{n} X_\mu X_\nu \cos[(\mu-\nu)\lambda]$$

$$= I(\lambda), \tag{2·2}$$

where $I(\lambda)$ is the periodogram intensity $G(\lambda) G^*(\lambda) \equiv A^2(\lambda) + B^2(\lambda)$, $G(\lambda)$ being calculated as

$$\sqrt{\frac{2}{n} \sum_{r=1}^{n} X_r e^{i\lambda r}} \tag{2·3}$$

(usually for integral p, where $\lambda = 2\pi p/n$). However, in spite of the relation (2·2), it is now known that $I(\lambda)$ does not provide a consistent estimate of $g(\lambda)$, owing to its large sampling fluctuations which do not diminish as n increases. The relevant stochastic properties of $I(\lambda)$ have been discussed by Slutsky (1927), Bartlett (1950, 1954), Grenander (1951) and others, and will merely be quoted when necessary. The further condition imposed on X_r for these properties to hold is that, if not normal, at least it has the linear form

$$X_t = \sum_{u=0}^{\infty} b_u Y_{t-u}, \tag{2·4}$$

where Y_t is a sequence of *independent* quantities with zero mean and constant variance σ^2.

In view of the sampling properties of $I(\lambda)$, 'smoothed' estimates of $g(\lambda)$ have been proposed. Thus Daniell put forward the estimate

$$g_D(\lambda) = \frac{1}{2h} \int_{\lambda-h}^{\lambda+h} I(\omega) \, d\omega; \tag{2·5}$$

and Bartlett's formula (cf. also Tukey and Hamming, 1949) for smoothing the periodogram $I(\lambda)$ is

$$g_B(\lambda) = 2 \sum_{s=-n'+1}^{n'-1} \left(1 - \frac{|s|}{n'}\right) C'_s \cos(s, \lambda), \tag{2·6}$$

where $C'_s = C_s \bigg/ \left(1 - \frac{|s|}{n}\right)$. The resolving power of this last procedure increases with n', and the smoothing with $m = n/n'$ (where n is the total length of the series), fluctuations being of order $1/\sqrt{m}$. Grenander subsequently examined a whole class of estimates containing the above two, by introducing a general weighting factor, $u_s(\lambda)$, in the formula

$$g_G(\lambda) = 2 \sum_{s=-n}^{n} u_s(\lambda) C_s \cos(s\lambda), \tag{2·7}$$

or in alternative form (for suitable conditions on $u_s(\lambda)$)

$$g_G(\lambda) = \int_0^\pi I(\omega) w_\lambda(\omega) \, d\omega, \tag{2·8}$$

where the positive weighting function $w_\lambda(\omega)$ is expressible in terms of $u_s(\lambda)$ by the formula

$$w_\lambda(\omega) = \frac{1}{\pi} \sum_{s=-n}^{n} u_s(\lambda) \cos(s\lambda) \cos(s\omega). \tag{2·9}$$

The estimate $g_G(\lambda)$ has the asymptotic sampling properties

$$E\{g_G(\lambda)\} \sim \int_0^\pi g(\omega) w_\lambda(\omega) \, d\omega, \tag{2·10}$$

$$\operatorname{var}\{g_G(\lambda)\} \sim \frac{2\pi}{n} \int_0^\pi g^2(\omega) w_\lambda^2(\omega) \, d\omega. \tag{2·11}$$

The relation of Daniell's estimate $g_D(\lambda)$ in (2·5) to the general formula (2·8) is obvious; the other estimate in (2·6) has for large n the effective weighting factor

$$u_s(\lambda) = \begin{cases} 1 - \dfrac{|s|}{n'} & (|s| < n'), \\ 0 & (|s| \geqslant n'), \end{cases} \tag{2·12}$$

and the weighting function corresponding to $u_s(\lambda)$ in (2·12) is

$$w_\lambda(\omega) = \frac{1}{2\pi n'} \left[\frac{\sin^2\{\tfrac{1}{2}n'(\omega - \lambda)\}}{\sin^2\{\tfrac{1}{2}(\omega - \lambda)\}} + \frac{\sin^2\{\tfrac{1}{2}n'(\omega + \lambda)\}}{\sin^2\{\tfrac{1}{2}(\omega + \lambda)\}} \right]. \tag{2·13}$$

Clearly

$$E\{g_D(\lambda)\} \sim \frac{1}{2h} \int_{\lambda - h}^{\lambda + h} g(\omega)\, d\omega$$

$$\sim g(\lambda) \quad \text{for small } h \text{ only,} \tag{2·14}$$

and

$$\operatorname{var}\{g_D(\lambda)\} \sim \frac{\pi}{nh} g^2(\lambda) \quad \text{for small } h. \tag{2·15}$$

Grenander has given also the results

$$E\{g_B(\lambda)\} \sim g(\lambda) \tag{2·16}$$

and

$$\operatorname{var}\{g_B(\lambda)\} \sim \begin{cases} \dfrac{2}{3m} g^2(\lambda) & (\lambda \neq 0), \\ \dfrac{4}{3m} g^2(0) & (\lambda = 0). \end{cases} \tag{2·17}$$

3. UNCERTAINTY RELATIONS

When Grenander investigated further the sampling properties of $g_G(\lambda)$, he imposed the condition

$$\int_0^\pi w_\lambda(\omega)\, d\omega = 1 \tag{3·1}$$

as a 'sort of asymptotic unbiasedness', but this condition (which is satisfied by the weighting function for $g_D(\lambda)$, even for h not small) appears unsatisfactory, and we shall use the correct condition for asymptotic unbiasedness from formula (2·10), viz.

$$\int_0^\pi g(w)\, w_\lambda(\omega)\, d\omega \sim g(\lambda). \tag{3·2}$$

With this condition the formula for $\operatorname{var}\{g_G(\lambda)\}$ in (2·11) yields

$$\frac{n}{2\pi} \operatorname{var}\{g_G(\lambda)\} \sim \int_0^\pi g^2(\omega)\, w_\lambda^2(\omega)\, d\omega \geqslant \frac{1}{\pi}\left[\int_0^\pi g(\omega)\, w_\lambda(\omega)\, d\omega \right]^2$$

$$\sim \frac{1}{\pi} g^2(\lambda),$$

whence asymptotically

$$\operatorname{var}\{g_G(\lambda)\} \geqslant \frac{2}{n} g^2(\lambda). \tag{3·3}$$

Moreover, the condition for equality is reached if

$$w_\lambda(\omega) = g(\lambda)/[\pi g(\omega)]. \tag{3·4}$$

10

This 'ideal' estimate thus has weighting function inversely proportional to $g(\omega)$,[†] not $g^2(\omega)$, as Grenander obtained with condition (3·1). However, it is still entirely theoretical in character, for it assumes a complete knowledge of the unknown spectral function $g(\omega)$. Furthermore, for different values of λ, the weighting function multiplying the stochastic quantity $I(\omega)$ remains inversely proportional to $g(\omega)$, and all the corresponding estimates $g_G(\lambda)$ are stochastically equivalent.

In an attempt to overcome such difficultues, Grenander made use of a measure of 'resolvability', defined in general by the 'standard deviation' of ω for the weighting function $w_\lambda(\omega)$. As $w_\lambda(\omega)$ no longer necessarily satisfies the 'normalization' condition (3·1), this measure is modified here to

$$U_1 = \left[\int_0^\pi (\omega - \lambda)^2 \, w_\lambda(\lambda) \, d\omega \Big/ \int_0^\pi w_\lambda(\omega) \, d\omega \right]^{\frac{1}{2}}, \tag{3·5}$$

(note also that the deviation $w - \lambda$ is used, whether or not the 'mean' ω is exactly λ). It is further generalized to

$$U_1^{(r)} = \left[\int_0^\pi | \omega - \lambda |^r \, w_\lambda(\omega) \, d\omega \Big/ \int_0^\pi w_\lambda(\omega) \, d\omega \right]^{1/r} \qquad (r > 0), \tag{3·6}$$

mainly in order to stress the rather arbitrary character of such measures. In terms of $U_1^{(r)}$ and

$$U_2 \equiv \int_0^\pi g^2(\omega) \, w_\lambda^2(\omega) \, d\omega \sim \frac{n}{2\pi} \operatorname{var}\{g_G(\lambda)\}, \tag{3·7}$$

we may now generalize Grenander's 'uncertainty principle', connecting $U_1^{(r)}$ and U_2 by the following modification of his argument.

From the Tchebychev inequality

$$Q\{| \omega - \lambda | < 2U_1^{(r)}\} > \left(1 - \frac{1}{2^r}\right),$$

where Q denotes 'probability' in terms of the formal frequency function

$$q(\omega) \equiv w_\lambda(\omega) \Big/ \int_0^\pi w_\lambda(\omega) \, d\omega,$$

$$\int_0^\pi q^2(\omega) \, d\omega \geqslant \int_{\lambda - 2U_1^{(r)}}^{\lambda + 2U_1^{(r)}} q^2(\omega) \, d\omega \geqslant \frac{1}{4U_1^{(r)}} \left[\int_{\lambda - 2U_1^{(r)}}^{\lambda + 2U_1^{(r)}} q(\omega) \, d\omega \right]^2$$

$$> \frac{1}{4U_1^{(r)}} \left[1 - \frac{1}{2^r} \right]^2.$$

Let A, B be the upper and lower bounds of $g(\omega)$ in $(0, \pi)$, these being assumed (respectively) finite and non-zero. Then

$$U_2 \geqslant B^2 \int_0^\pi w_\lambda^2(\omega) \, d\omega$$

$$> \frac{B^2 k_0^2}{4U_1^{(r)}} \left[1 - \frac{1}{2^r} \right]^2,$$

where

$$k_r \equiv \int_0^\pi \omega^r w_\lambda(\omega) \, d\omega. \tag{3·8}$$

[†] The ideal weighting of $I(\omega)$ by a function varying as $1/g(\omega)$ is in entire accord with the asymptotic transformation to a uniform spectrum often desirable in spectral analysis (see, for example, Bartlett, 1954).

Now
$$Ak_0 > \int_0^\pi w_\lambda(\omega) g(\omega) \, d\omega \sim g(\lambda),$$

whence finally
$$U \equiv U_1^{(r)} U_2 > \tfrac{1}{4} g^2(\lambda) \frac{|B^2}{|A^2} \left| 1 - \frac{1}{2^r} \right|^2 > 0. \tag{3.9}$$

which shows that the product U of the two quantities $U_1^{(r)}$ and U_2 has a positive minimum.

The condition of asymptotic unbiasedness which we have imposed could perhaps be waived by considering the mean-square deviation of $g_G(\lambda)$ from $g(\lambda)$: this would, however, no longer give a criterion like U_2 asymptotically independent of n, so that it seems more convenient to adopt the conditions above.

4. ASYMPTOTIC EFFICIENCIES OF $g_D(\lambda)$ AND $g_B(\lambda)$

From (3.9) we may define a quantity
$$W \equiv U / g^2(\lambda) \tag{4.1}$$

as a measure of uncertainty depending on the form of the estimate $g_G(\lambda)$, and on the choice of r in $U_1^{(r)}$. We compare $g_D(\lambda)$ and $g_B(\lambda)$ by this criterion. The 'ideal' estimate (for $r = 2$) with the weighting function given in (3.4) is modified to
$$w_\lambda(\omega) = g(\lambda)/[2hg(\omega)] \tag{4.2}$$

if restricted to the interval $(\lambda - h, \lambda + h)$. For small h, $g(\omega) \sim g(\lambda)$, so that $g_D(\lambda)$ approximates to the 'ideal' estimate *if the latter is so restricted*. As, for $g_D(\lambda)$ with small h,
$$U_1^{(r)} \sim h(r+1)^{-1/r}, \quad U_2 \sim \tfrac{1}{2} g^2(\lambda)/h, \tag{4.3}$$

we have
$$W \sim \tfrac{1}{2}(r+1)^{-1/r}. \tag{4.4}$$

Table 1

	$\lambda = \tfrac{1}{4}\pi$		$\lambda = \tfrac{1}{2}\pi$	
h:	$\tfrac{1}{30}\pi$	0·290	$\tfrac{1}{16}\pi$	0·293
	$\tfrac{1}{15}\pi$	0·292	$\tfrac{1}{8}\pi$	0·306
	$\tfrac{1}{10}\pi$	0·301	$\tfrac{3}{16}\pi$	0·314
	$\tfrac{2}{15}\pi$	0·311	$\tfrac{3}{8}\pi$	0·321
	$\tfrac{1}{6}\pi$	0·322	$\tfrac{1}{4}\pi$	0·322

A numerical calculation (Table 1) was made of W for the restricted 'ideal' estimate ($r = 2$) for varying h in the particular case of the spectrum of M. G. Kendall's artificial series I (Kendall, 1946). As, for small h, $W \to 0.289$ (for all λ *within* the total range), the above results show that W is smallest, in this particular case at least, as $h \to 0$, for the above class of estimate. More generally, for small h, it is readily shown that
$$W = \frac{1}{2\sqrt{3}} \left(1 + \frac{2ch^2}{15} \right), \tag{4.5}$$

where
$$c = \left(\frac{1}{g(\lambda)} \frac{dg(\lambda)}{d\lambda} \right)^2 - \frac{1}{2g(\lambda)} \frac{d^2 g(\lambda)}{d\lambda^2}, \tag{4.6}$$

so that whether W increases or decreases as h increases from zero depends on the sign of c; it will, however, increase in the neighbourhood of a maximum of $g(\lambda)$.

For $g_B(\lambda)$ it can be shown (e.g. by the use of (2·13)) that

$$[l_1^{(2)}]^2 \sim \frac{4}{n'}\left[\left(\frac{\pi-\lambda}{\pi}\right)\log\cos\tfrac{1}{2}\lambda + \frac{\lambda}{\pi}\log\sin\tfrac{1}{2}\lambda + \frac{1}{\pi}\log 2\right].$$

$$l_1^{(1)} \sim \frac{2}{\pi n'}\log n'. \tag{4·7}$$

$$[U_1^{(r)}]^r \sim 1/[(n')^r \cos(\tfrac{1}{2}\pi r)\,\Gamma(2-r)] \quad (0 < r < 1).$$

Making use of the result (2·17) (for $\lambda \neq 0$), we arrive at the comparison given in Table 2 of $g_D(\lambda)$ and $g_B(\lambda)$ in terms of W. Asymptotically, $g_D(\lambda)$ is thus superior to $g_B(\lambda)$ if the criterion W is taken for $r = 2$ (or 1), but the generalization of the resolvability measure U_1 emphasized its arbitrary character, and the 'efficiencies' of the two estimates cross over as r decreases.

Table 2

r	$g_D(\lambda)$	$g_B(\lambda)$
2	0·289	0·125 $\sqrt{n'}$ $(\lambda = \tfrac{1}{2}\pi)$
1	0·250	0·0675 log n'
$\tfrac{1}{2}$	0·222	0·270
$\tfrac{1}{4}$	0·211	0·222
0·1	0·193	0·177
0+	0·184	0·162

5. INVESTIGATION OF THE THEORETICAL OPTIMUM WEIGHTING FUNCTION

The replacement of U_2 by $U_1^{(r)}U_2$ as a criterion of efficiency naturally complicates the investigation of the 'ideal' weighting function, but while this will still of course depend on a knowledge of the theoretical spectrum, it seemed worth while investigating the form of the optimum function if this criterion were adopted. We examine the case $r = 2$. Let $w_\lambda(\omega) \equiv v^2(\omega)$, where $v(\omega)$ is real, be the optimum function, and consider the slightly modified function $v(\omega) + \eta V(\omega)$, where η is small and $V(\omega)$ arbitrary. Denote the change in $U_1 U_2$ by $\delta(U_1 U_2)$. Now from (3·8)

$$l_1 = [(k_2 - 2\lambda k_1 + \lambda^2 k_0)\cdot k_0]^{\frac{1}{2}}. \tag{5·1}$$

and, to the first order of η,

$$\delta(k_n) = 2\eta K_n,$$

where

$$K_n \equiv \int_0^\pi \omega^n v(\omega)\,V(\omega)\,d\omega. \tag{5·2}$$

Hence

$$\delta(l_1) = \eta\{(K_2 - 2\lambda K_1 - \lambda^2 K_0)(l_1 k_0) - K_0 l_1 k_0\}.$$

Also

$$\delta(l_2) = 4\eta\int_0^\pi g^2(\omega)\,v^3(\omega)\,V(\omega)\,d\omega = 4\eta L, \text{ say.}$$

Thus finally

$$\delta(l_1 l_2) = \frac{\eta}{k_0}\left[\frac{(K_2 - 2\lambda K_1 + \lambda^2 K_0)l_2}{l_1} - K_0 l_1 l_2 + 4k_0 l_1 L\right]. \tag{5·3}$$

As $v(\omega)$ is assumed to be the optimum function, the coefficient of η should vanish. This implies that the expression within the square brackets in (5·3) should vanish for any appropriate $V(\omega)$. The condition of asymptotic unbiasedness gives the condition

$$\int_0^\pi g(\omega)\,v(\omega)\,V(\omega)\,d\omega = 0, \tag{5·4}$$

which will be satisfied if we take in particular

$$V(\omega)\,d\omega \equiv d\left|\frac{\epsilon(\omega-\alpha)}{g(\alpha)\,v(\alpha)} - \frac{\epsilon(\omega-\beta)}{g(\beta)\,v(\beta)}\right|,$$

where

$$\epsilon(x) \equiv \begin{cases} 0 & (x<0), \\ 1 & (x\geqslant 0). \end{cases}$$

The expression in square brackets in (5·3) now becomes

$$\frac{U_2}{U_1}\left[\frac{(\alpha-\lambda)^2}{g(\alpha)} - \frac{(\beta-\lambda)^2}{g(\beta)}\right] - U_1 U_2\left[\frac{1}{g(\alpha)} - \frac{1}{g(\beta)}\right] + 4k_0 U_1[g(\alpha)\,v^2(\alpha) - g(\beta)\,v^2(\beta)].$$

and, as this should be zero for arbitrary α, β, we obtain

$$\frac{U_2}{U_1}\frac{(\alpha-\lambda)^2}{g(\alpha)} - \frac{U_1 U_2}{g(\alpha)} + 4k_0 U_1 g(\alpha)\,v^2(\alpha) = \text{constant},$$

or

$$v^2(\alpha) = \frac{A}{g(\alpha)} + \frac{B}{g^2(\alpha)}\left[1 - \frac{(\alpha-\lambda)^2}{U_1^2}\right], \tag{5·5}$$

which determines the optimum form (if this exists in the above sense). It depends of course, as anticipated, on the unknown theoretical spectral function $g(\omega)$. In (5·5),

$$B = \tfrac{1}{2}U_2/k_0, \tag{5·6}$$

and the two formulae for U_1 and U_2 give, with the unbiasedness condition, three further relations for A, B, U_1 and U_2. If these yield an appropriate solution, the minimum $U_1 U_2$ is then also calculable.

6. Equations for the optimum function in the case of the second-order autoregressive process

The determination of the coefficients in (5·5) is unfortunately extremely laborious, and sometimes impossible, as was experienced in the particular case of the second-order autoregressive process

$$X_{t+2} + aX_{t+1} + bX_t = Y_{t+2}. \tag{6·1}$$

For (6·1) we have

$$g(\lambda) = 2\pi\sigma_X^2 f(\lambda) = \delta/(\alpha + \beta\cos\lambda + \gamma\cos 2\lambda), \tag{6·2}$$

where $\alpha = 1 + a^2 + b^2$, $\beta = 2a(1+b)$, $\gamma = 2b$, $\delta = 2\sigma^2$ (see Bartlett, 1950). Define

$$a_n^{(m)} = \int_0^\pi \omega^m g^{-n}(\omega)\,d\omega, \quad b_n^{(m)} = \int_0^\pi (\omega-\lambda)^m g^{-n}(\omega)\,d\omega,$$

which are evaluable for the $g(\lambda)$ of (6·2). Then it will be found that (5·6) yields the equation

$$\pi A^2 + 2A(a_1^{(0)}B - b_1^{(2)}C) + a_2^{(0)}B^2 - 2b_2^{(2)}BC + b_2^{(4)}C^2 = 4Bk_0, \tag{6·3}$$

where $C = B/U_1^2$, and

$$Aa_1^{(0)} + Ba_2^{(0)} - Cb_2^{(2)} = k_0. \tag{6·4}$$

The relation for U_1 give

$$Ab_1^{(2)} + Bb_2^{(2)} - Cb_2^{(4)} = k_0 B/C, \tag{6·5}$$

and the unbiasedness condition gives

$$A\pi + Ba_1^{(0)} - Cb_1^{(2)} = g(\lambda). \tag{6·6}$$

Substitution of k_0 in (6·3) and (6·5) gives, with the aid of (6·6), two equations of second degree in A and B. A being found in terms of B from one of them, a biquadratic equation in

A results from the other. An examination of an actual numerical case of (6·2) revealed that no solution in the admissible ranges for the unknowns is always possible; in any case, if we recall that the solutions even when they exist are only available when $g(\lambda)$ is known, these equations seem unlikely to be used in practice.

7. Theoretical gain in efficiency with weighting function in restricted range

In view of the difficulty of handling the optimum weighting function (on the basis of the criterion U), it may be advisable to demonstrate that the optimum function on the basis of the criterion U_2 but restricted to the *small* interval $\lambda - h$, $\lambda + h$, viz.

$$w(\lambda) = g(\lambda)/[2hg(\omega)] \sim 1/2h, \tag{7.1}$$

may at least be improved on the basis of the criterion U. As, for small h, $g(\omega) \sim g(\lambda)$ to the first order, the weighting function in (5·5) may be written to the same degree of approximation

$$A' + B'(\omega - \lambda)^2,$$

and it is easily found from the further relations for A, B, U_1 and U_2 that

$$w_W(\omega) \sim \frac{3}{4h}\left(1 - \frac{(\omega - \lambda)^2}{h^2}\right); \tag{7.2}$$

Moreover, for (7·2), $W = 3\sqrt{5}/25 = 0.268.$

This is only a small improvement over the value 0·289 for W with the Daniell estimate $g_D(\lambda)$, but it will be noticed that this approximate weighting function is independent of the unknown spectral function $g(\lambda)$, and hence also usable in practice. As the approximate bias in the resulting estimate $g_W(\lambda)$ is reduced from $\frac{1}{6}h^2 d^2 g(\lambda)/d\lambda^2$ to $\frac{1}{10}h^2 d^2 g(\lambda)/d\lambda^2$, the use of $g_W(\lambda)$ in place of $g_D(\lambda)$ might sometimes be preferable when the periodogram $I(\omega)$ is available.

The merits of the other estimate considered, $g_B(\lambda)$, rest in its asymptotic unbiasedness and its convenience of calculation from the first $n' - 1$ sample autocovariances (or autocorrelations). From Table 2 we saw that its efficiency depended on what measure of resolvability was used.

REFERENCES

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. *J. Roy. Statist. Soc. Suppl.* **8**, 27.
Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika*, **37**, 1.
Bartlett, M. S. (1954). Problèmes de l'analyse spectrale des séries temporelles stationnaires. *Publ. de Inst. Statist. (Univ. de Paris)* vol. III, fasc. 3, 119.
Grenander, U. (1951). On empirical spectral analysis of stochastic processes. *Ark. Mat.* **1**, 503.
Grenander, U. & Rosenblatt, M. (1952). On spectral analysis of stationary time-series. *Proc. Nat. Acad. Sci., Wash.*, **38**, 519.
Grenander, U. & Rosenblatt, M. (1954). Statistical spectral analysis of time-series arising from stationary stochastic processes. *Ann. Math. Statist.* **24**, 537.
Kendall, M. G. (1946). *Oscillatory Time Series.* Cambridge University Press.
Slutsky, E. (1927). The summation of random causes as the source of cyclic processes. Problems of economic conditions, ed. by the Conjuncture Institute, Moscow, **3**, no. 1. (Later reprinted in *Econometrica*, **5** (1937), 105.)
Tukey, J. W. & Hamming, R. W. (1949). Measuring noise color, I. *Bell Tel. Lab. Memo.* MM–49–110.
Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*, 1st ed. Uppsala.

THE AUTOCORRELATION FUNCTION AND THE SPECTRAL DENSITY FUNCTION

By J. WISE

London School of Economics

INTRODUCTION AND SUMMARY

The relationship existing between the autocorrelation function and the spectral density function of a stationary, purely non-deterministic stochastic process is well known. Discussions of the relationship have been given by Wiener (1930) and Khintchine (1934), in the case of the continuous process, and by Wold (1954, Chapter 2, § 17, pp. 66–75), and Doob (1953, Chapter 10, §§ 3 and 4, pp. 473–86) for the discrete process. The treatment in the present paper is confined to the discrete process.*

The relationship referred to also provides a fundamental connexion between the autocorrelation matrix of a process and the corresponding spectral density function of the process. Some rather ingenious manipulations on the matrix representation of the spectral function have been carried out by Whittle (1951, Chapter 4, § 2, pp. 35–6, equations (4·272) and (4·276)). However, Whittle's work contains some inaccuracies. More important still, 'exact' results exist for those cases in which Whittle gave 'asymptotic' solutions.

In the course of the derivations of these exact relationships, the author uses methods which appear to be new—though bearing some relationship to André's method of solving linear difference equations (see C. Jordan, 1947, pp. 587–99)—and are capable of wide application in the theory of discrete linear stochastic processes. The use of exact relationships does, moreover, lead to a considerable improvement in clarity and rigour.

In view of the fundamental exact relationships existing between the autocovariance matrix and the spectral density function, the former might appropriately—in the case of the stationary process—be termed the spectral density matrix of the process. Traditional terminology has, however, been adhered to in the remainder of the paper.

The main topics treated below may be classified under the following headings.

(a) The circular process

(i) The exact relationship existing between the autocovariance matrix and the spectral density function.

(ii) The latent roots of the autocovariance matrix.

(iii) The exact relationship between the canonical form of the autocovariance matrix and the spectral density function.

(b) The non-circular process

(i) The exact relationship corresponding to (a) (i) above.

(ii) The exact inversion of non-circular autocovariance matrices.

* This discrete process may, however, be a sample realization drawn from a continuous process.

Definitions

The vector of N random variables, given by $\mathbf{x}' = \{x_N, x_{N-1}, ..., x_1\}$, is assumed to have the following distributional properties:

$$E\mathbf{x} = 0, \tag{1}$$

$$Ex_s^2 = \sigma^2 \quad \{s = 1, 2, ..., N\}, \tag{2}$$

$$Ex_s x_{s+L} = \sigma^2 \rho_L \quad \{s = 1, 2, ..., N\}, \tag{3}$$

where
$$\rho_{N+L} = \rho_{N-L} = \rho_L.$$

The sequence of values $\rho_1, \rho_2, ..., \rho_N$ is termed the autocorrelation function of the process.

$$E\mathbf{xx}' = \mathbf{V} \quad \text{(the autocovariance matrix)}$$

$$= \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \cdots & \rho_2 & \rho_1 \\ \rho_1 & 1 & \rho_1 & \rho_2 & \cdots & \rho_3 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 & \cdots & \rho_4 & \rho_3 \\ \vdots & & & & & & \vdots \\ \rho_1 & \rho_2 & \rho_3 & \rho_4 & \cdots & \rho_1 & 1 \end{bmatrix}, \tag{4}$$

an $N \times N$ matrix.

Properties

The autocovariance matrix, $\mathbf{V}$, may, from (4), be expressed in the following form:

$$\mathbf{V} = \sigma^2 \{\mathbf{I} + \rho_1(\mathbf{W} + \mathbf{W}^{-1}) + \rho_2(\mathbf{W}^2 + \mathbf{W}^{-2}) + ... + \rho_{\frac{1}{2}(N-1)}(\mathbf{W}^{\frac{1}{2}(N-1)} + \mathbf{W}^{-\frac{1}{2}(N-1)})\}, \tag{5}$$

where N is a positive odd integer, or, alternatively,

$$= \sigma^2 \{\mathbf{I} + \rho_1(\mathbf{W} + \mathbf{W}^{-1}) + \rho_2(\mathbf{W}^2 + \mathbf{W}^{-2}) + ... + \tfrac{1}{2}\rho_{\frac{1}{2}N}(\mathbf{W}^{\frac{1}{2}N} + \mathbf{W}^{-\frac{1}{2}N})\},$$

where N is a positive even integer. For all values of N, $\mathbf{I}$ denotes the $N \times N$ identity matrix, and $\mathbf{W}$ denotes the $N \times N$ circulant definition of the auxiliary identity matrix, that is to say,

$$\mathbf{W} = 0 \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & 0 & . & . & \cdots & 1 \\ 1 & 0 & . & . & \cdots & 0 \end{bmatrix}. \tag{6}$$

Since the matrices $(\mathbf{W}^L + \mathbf{W}^{-L})$ for $(L = 1, 2, ..., N)$ are commutative, they may all be simultaneously reduced to canonical form by a single transformation.

Thus, if we consider the matrix $\mathbf{L}$, satisfying the property

$$\mathbf{L}'\mathbf{L} = \mathbf{I}, \tag{7}$$

$$\mathbf{L}'(\mathbf{W}^L + \mathbf{W}^{-L})\mathbf{L} = \text{diag}\{\omega_1^L + \omega_1^{-L}, ..., \omega_N^L + \omega_N^{-L}\}, \tag{8}$$

where $\omega_1, \omega_2, ..., \omega_N$ are the N values of the roots of the equation

$$y^N - 1 = 0, \tag{9}$$

then
$$\mathbf{L}'\mathbf{VL} = \text{diag}\{\nu_1, \nu_2, \nu_3, ..., \nu_N\}$$

provides the canonical form of the autocovariance matrix $\mathbf{V}$.

Furthermore, we have

$$\omega_s^L + \omega_s^{-L} = 2\cos\frac{2\pi sL}{N} \qquad (s = 1, 2, \ldots, N, \; L = 1, 2, \ldots, N). \tag{10}$$

The latent roots of V are thus given by

$$\nu_s = \sigma^2\{1 + \rho_1(\omega_s + \omega_s^{-1}) + \rho_2(\omega_s^2 + \omega_s^{-2}) + \ldots + \rho_{\frac12(N-1)}(\omega_s^{\frac12(N-1)} + \omega_s^{-\frac12(N-1)})\} \qquad (s = 1, 2, \ldots, N)$$

$$= \sigma^2\left\{1 + 2\rho_1\cos\frac{2\pi s}{N} + 2\rho_2\cos\frac{4\pi s}{N} + \ldots + 2\rho_{\frac12(N-1)}\cos\frac{(N-1)\pi s}{N}\right\} \qquad (s = 1, 2, \ldots, N)$$

for N odd, or by

$$\nu_s = \sigma^2\{1 + \rho_1(\omega_s + \omega_s^{-1}) + \rho_2(\omega_s^2 + \omega_s^{-2}) + \ldots + \tfrac12\rho_{\frac12N}(\omega_s^{\frac12N} + \omega_s^{-\frac12N})\} \qquad (s = 1, 2, \ldots, N)$$

$$= \sigma^2\left\{1 + 2\rho_1\cos\frac{2\pi s}{N} + 2\rho_2\cos\frac{4\pi s}{N} + \ldots + \rho_{\frac12N}\cos\pi s\right\} \qquad (s = 1, 2, \ldots, N) \tag{11}$$

for N even.

The relationships given by (5) and (11) are the correct restatement of similar forms given by Whittle (1951, Chapter 4, §2, pp. 35–6), which are incorrect even as approximations in the non-circular case.

General rules for the multiplicities of the latent roots of V have been incorrectly stated by both Whittle (1951, Chapter 4, §3, p. 37) and Watson (1951*, Chapter 2, §5, p. 50).

Results correctly obtained by R. L. Anderson (1942, §5a, pp. 7–8), on the multiplicities in the values of the terms in the sequence,

$$\cos\frac{2\pi L}{N}, \quad \cos\frac{4\pi L}{N}, \quad \cos\frac{6\pi L}{N}, \quad \ldots, \quad \cos 2\pi L, \tag{12}$$

may, however, be used to derive valid statements relating to the multiplicities in the latent roots of V.

Let $N = nk$ and $L = lk$, where n, l and k are positive integers, n and l being prime to one another. Two cases may then be distinguished.

(i) For n odd, there are $\frac12(n-1)$ terms in the series (12) occurring with $2k$-fold multiplicities, together with the value unity occurring with k-fold multiplicity.

(ii) For n even, there are $\frac12(n-2)$ terms in the series (12) occurring with $2k$-fold multiplicities, together with each of the values ± 1 occurring with k-fold multiplicities. For the derivation of these results the reader is referred to R. L. Anderson's paper.

Although it is impossible to enumerate here all the possibilities in detail, a 'rule' which the author has found useful in applications of the theory to sampling problems will be given relating to the multiplicities in the latent roots of V. This 'rule' is adequate for most of the cases which occur in practice. The 'rule' is not, however, a theorem, since exceptions to it occur when certain relationships exist between the elements of V. Care must therefore be taken when it is applied. This 'rule' is as follows:

If $\rho_L \neq 0$ for at least one value of L satisfying the condition $N = nk$ and $L = lk$, n and l being prime to one another, then the multiplicities in the latent roots will be at most $2k$-fold.

Whittle's assertion that if N is odd there are $\frac12(N-1)$ pairs of distinct roots, with a single root greater than or less than all of these others; and if N is even there are $\frac12(N-2)$ pairs of

* Watson states here that 'real symmetric circulant matrices have at most one odd latent root, the rest being pairwise equal'. That is, of course, incorrect.

distinct roots with one root greater than and one root less than all of the other roots, is incorrect.

An example occurring in the literature on time-series analysis, demonstrating a practical instance in which the 'rule' stated above yields the correct result, whereas Whittle's assertion breaks down, is the lag $- L$ definition of the Markoff process, given by

$$x_t - \rho x_{t-L} = \epsilon_t \quad (t = 1, 2, ..., N), \tag{13}$$

where $x_{N+S} = x_S$, N and L not being prime to one another.

The spectral function of a circular process

The spectral density function of the circular process possessing the distributional properties (1), (2) and (3) is of the form*

$$\nu(\theta) = \sigma^2 \left| 1 + 2\rho_1 \cos\theta + 2\rho_2 \cos 2\theta + ... + 2\rho_{\frac{1}{2}(N-1)} \cos \frac{(N-1)\theta}{2} \right| \quad \left| \theta = \frac{2\pi}{N}, \frac{4\pi}{N}, ..., 2\pi \right|,$$

where N is odd; or of the form

$$\nu(\theta) = \sigma^2 \{1 + 2\rho_1 \cos\theta + 2\rho_2 \cos 2\theta + ... + \rho_{\frac{1}{2}N} \cos \tfrac{1}{2}N\theta\} \quad \left\{\theta = \frac{2\pi}{N}, \frac{4\pi}{N}, ..., 2\pi \right\}, \tag{14}$$

where N is even.

The latent roots of $\mathbf{V}$, given by (11), are thus the values of the spectral density function, $\nu(\theta)$, taken at constant intervals of $2\pi/N$ in the values of θ.

These roots may thus be regarded as being generated by the spectral density function of the process.

The results obtained so far are, of course, exact for finite N. By considering the limiting values of the roots as N tends to infinity, asymptotic results may be obtained.

The autoregressive-moving average process

The discrete circular linear process of the autoregressive-moving average type may be written in the form

$$x_t + \alpha_1 x_{t-1} + ... + \alpha_k x_{t-k} = \epsilon_t + \beta_1 \epsilon_{t-1} + ... + \beta_h \epsilon_{t-h} \quad \{t = 1, 2, ..., N\}, \tag{15}$$

where $x_{N+S} = x_S$, $\epsilon_{N+S} = \epsilon_S$, and

$$\begin{array}{ll} E\epsilon_s = 0 & \{s = 1, 2, ..., N\}, \\ E\epsilon_s^2 = \sigma^2 & \{s = 1, 2, ..., N\}, \\ E\epsilon_s \epsilon_t = 0 & \{s \neq t\}. \end{array} \tag{16}$$

---

* Any discrete stationary purely non-deterministic process may be written in the form

$$x_t = \eta_t + \beta_1 \eta_{t-1} + \beta_2 \eta_{t-2} + ...,$$

where the summation extends to infinity.

The spectral density function of this process is defined as

$$\nu(\theta) = (1 + \beta_1 e^{i\theta} + \beta_2 e^{2i\theta} + ...)(1 + \beta_1 e^{-i\theta} + \beta_2 e^{-2i\theta} + ...) \text{var } \eta$$

for both circular and non-circular definitions of the process, for samples consisting of any number of consecutive observations. The relation (14) is obtained by expanding $\nu(\theta)$, thus defined, as a Fourier series.

In the matrix notation the process becomes

$$(I + \alpha_1 W + \ldots + \alpha_k W^k) x = (I + \beta_1 W + \ldots + \beta_h W^h) \epsilon. \tag{17}$$

where

$$x' = \{x_N, x_{N-1}, x_{N-2}, \ldots, x_2, x_1\}$$

and

$$\epsilon' = \{\epsilon_N, \epsilon_{N-1}, \epsilon_{N-2}, \ldots, \epsilon_2, \epsilon_1\}.$$

Solving (17) for x, we obtain

$$x = (I + \alpha_1 W + \ldots + \alpha_k W^k)^{-1} (I + \beta_1 W + \ldots + \beta_h W^h) \epsilon. \tag{18}$$

Thus, from (18), we obtain

$$Ex = 0, \tag{19}$$

and

$$Exx' = V \quad \text{(the autocovariance matrix)}$$

$$= \sigma^2 (I + \alpha_1 W + \ldots + \alpha_k W^k)^{-1} (I + \beta_1 W + \ldots + \beta_h W^h)(I + \beta_1 W^{-1} + \ldots + \beta_h W^{-h})$$

$$\times (I + \alpha_1 W^{-1} + \ldots + \alpha_k W^{-k})^{-1}. \tag{20}$$

The latent roots of V are therefore given by

$$v_s = \sigma^2 \frac{(1 + \beta_1 \omega_s + \ldots + \beta_h \omega_s^h)(1 + \beta_1 \omega_s^{-1} + \ldots + \beta_h \omega_s^{-h})}{(1 + \alpha_1 \omega_s + \ldots + \alpha_k \omega_s^k)(1 + \alpha_1 \omega_s^{-1} + \ldots + \alpha_k \omega_s^{-k})} \quad (s = 1, 2, \ldots, N)$$

$$= \sigma^2 \frac{(1 + \beta_1 e^{2i\pi s/N} + \ldots + \beta_h e^{2i\pi hs/N})(1 + \beta_1 e^{-2i\pi s/N} + \ldots + \beta_h e^{-2i\pi hs/N})}{(1 + \alpha_1 e^{2i\pi s/N} + \ldots + \alpha_k e^{2i\pi ks/N})(1 + \alpha_1 e^{-2i\pi s/N} + \ldots + \alpha_k e^{-2i\pi ks/N})} \quad (s = 1, 2, \ldots, N). \tag{21}$$

The spectral density function of the circulant process (15) may be written as

$$v(\theta) = \sigma^2 \frac{(1 + \beta_1 e^{i\theta} + \ldots + \beta_h e^{ih\theta})(1 + \beta_1 e^{-i\theta} + \ldots + \beta_h e^{-ih\theta})}{(1 + \alpha_1 e^{i\theta} + \ldots + \alpha_k e^{ik\theta})(1 + \alpha_1 e^{-i\theta} + \ldots + \alpha_k e^{-ik\theta})} \quad (0 \leqslant \theta \leqslant 2\pi) \tag{22}$$

$$= g(e^{i\theta}) \quad \text{(let us say)} \quad (0 \leqslant \theta \leqslant 2\pi). \tag{23}$$

It follows from (20) and (22) that

$$V = g(W), \tag{24}$$

and, in addition,

$$v_s = g(\omega_s) = g(e^{2i\pi s/N}) \quad (s = 1, 2, \ldots, N).$$

These results are exact, and are fundamental to the exact treatment of circular processes.

An asymptotic relationship between V (in the non-circular case, to be further discussed in § (b) of this paper) and g(W) has been given by Whittle (1951). It is not at all clear in what sense this asymptotic relationship is valid, unless the exact form is used. Whittle (1952, Theorem 1, (6) and (7), pp. 49–50), and Wold (1953, Chapter 11, § 2, Theorem 1) use the following interpretation of the asymptotic relationship:

$$\lim_{N \to \infty} \frac{x' V x}{x' g(W) x} = 1 \tag{25}$$

for almost all realizations of the process. This result, though valid, is difficult to work with rigorously. Whittle does not adhere to it in any of his derivations, and for a rigorous treatment it is preferable to work throughout with exact relationships. The latter are no more cumbersome than the asymptotic forms, and their use may lead to the avoidance of substantial errors.

This is abundantly illustrated in the case of the first-order autoregressive process in the circular case. The largest latent root of the autocorrelation matrix for $x_1 < 0$ is given by $\dfrac{1-(-\alpha_1)^N}{1+(-\alpha_1)^N}\dfrac{1-\alpha_1}{1+\alpha_1}$. The corresponding asymptotic form of this is $\dfrac{1-\alpha_1}{1+\alpha_1}$, which differs substantially from the exact value when N is small and $-\alpha_1$ is close to unity. In fact, as $-\alpha_1$ tends to unity, the exact form tends in value to N, while the asymptotic form tends to infinity.

A final point which should be made on the subject of the latent roots of circulant variance matrices is that in the derivation of the values of, and of the multiplicities in these roots, it is usually more convenient to consider the spectral density function in the form (22) than in the form (14).

(b) THE NON-CIRCULAR PROCESS

Whittle (1951, Chapter 4, § 2, relations (4·253), (4·262) and (4·28), pp. 34–6) used the 'asymptotic' form of (23) to obtain a simple method for the 'approximate' inversion of **V** in the non-circular case. For many purposes, approximate inversion is not adequate, and is not in any case necessary. It will be shown below that a very simple method of exact inversion is available. Whittle (1951, Chapter 4, § 2, relations (4·242)–(4·252), pp. 33–4) does give a cumbersome method of exact inversion based on the auxiliary identity matrix, denoted by **U**, where

$$
\mathbf{U} = \begin{bmatrix}
0 & 1 & 0 & 0 & \dots & 0 \\
0 & 0 & 1 & 0 & \dots & 0 \\
0 & 0 & 0 & 1 & \dots & 0 \\
\vdots & & & & \ddots & \vdots \\
0 & 0 & 0 & 0 & \dots & 1 \\
0 & 0 & 0 & 0 & \dots & 0
\end{bmatrix}
\tag{26}
$$

an $N \times N$ matrix. Unfortunately Whittle's method appears to give the required result only in the case of the first-order autoregressive process.

An exact method of inversion will be given below which utilizes a non-circular counterpart of (20), the relation between the autocovariance matrix and the spectral density function.

The non-circular discrete linear process, generated by the relationship

$$
x_l + \alpha_1 x_{l-1} + \dots + \alpha_k x_{l-k} = \epsilon_l + \beta_1 \epsilon_{l-1} + \dots + \beta_h \epsilon_{l-h},
\tag{27}
$$

will be assumed to be semi-infinite. This does not prevent us from deriving results from samples consisting of N successive terms of the process. Such results for finite samples follow immediately from the case of the semi-infinite process. This is due to the fact that the autocovariance function and the spectral density function are both completely independent of N, a property which does not hold in the circular case.

We will assume (27) to represent a purely non-deterministic stationary process. An important theorem obtained by Wold (1954, Chapter 2, § 20, pp. 84–9, especially Theorem 7, p. 89) states, furthermore, that every semi-infinite, stationary, purely non-deterministic process can be expressed in the form (27).

A condition which is both necessary and sufficient for the stationarity of the process (27) can be obtained from a result given by Doob (1953).

The spectral density function of the process (27) is given by

$$g(e^{i\theta}) = \sigma^2 \frac{(1+\beta_1 e^{i\theta}+\ldots+\beta_h e^{ih\theta})(1+\beta_1 e^{-i\theta}+\ldots+\beta_h e^{-ih\theta})}{(1+\alpha_1 e^{i\theta}+\ldots+\alpha_k e^{ik\theta})(1+\alpha_1 e^{-i\theta}+\ldots+\alpha_k e^{-ik\theta})} \quad (0 \leqslant \theta \leqslant 2\pi).$$

It may, in passing, be noted that this is the same as (22), given for the circular process.

According to Doob (1953, Chapter 10, pp. 452-506, especially §10, pp. 501-6), the necessary and sufficient condition for a stochastic process with the spectral function (22) to be stationary in the 'wide sense' (Chapter 2, §8, pp. 94-5) is that the equation

$$z^h + \beta_1 z^{h-1} + \ldots + \beta_h = 0$$

shall have no roots outside the unit circle, and that the roots of the equation

$$z^k + \alpha_1 z^{k-1} + \ldots + \alpha_k = 0$$

shall all lie inside the unit circle. The reader is referred to Doob's* penetrating exposition for a more detailed treatment of this topic.

The relationship (27) may be transcribed as

$$(I+\alpha_1 U+\ldots+\alpha_k U^k)x = (I+\beta_1 U+\ldots+\beta_h U^h)\epsilon, \tag{28}$$

where x and ϵ are semi-infinite vectors, and U is the semi-infinite auxiliary identity matrix.

Solving (28) for x, leads to

$$x = (I+\alpha_1 U+\ldots+\alpha_k U^k)^{-1}(I+\beta_1 U+\ldots+\beta_h U^h)\epsilon. \tag{29}$$

From (29) we deduce

$$Ex = 0, \tag{30}$$

and

$$\begin{aligned}
V = Exx' \\
= \sigma^2(I+\alpha_1 U+\ldots+\alpha_k U^k)^{-1}(I+\beta_1 U+\ldots+\beta_h U^h)(I+\beta_1 U'+\ldots+\beta_h U'^h) \\
\times (I+\alpha_1 U'+\ldots+\alpha_k U'^k)^{-1},
\end{aligned} \tag{31}$$

which is thus the autocovariance matrix of the process (27). This is the exact, non-circular counterpart to the relationships (20) and (23), given above for the case of the circular process.

Relationship (31) bears a close affinity to the 'autocovariance generating function' given by

$$g(z) = \sigma^2 \frac{(1+\beta_1 z+\ldots+\beta_h z^h)(1+\beta_1 z^{-1}+\ldots+\beta_h z^{-h})}{(1+\alpha_1 z+\ldots+\alpha_k z^k)(1+\alpha_1 z^{-1}+\ldots+\alpha_k z^{-k})}. \tag{32}$$

However, (31) is much more useful. From it can be obtained directly (i) the autocovariance function of the process (27), this being the property (31) shares with (32): (ii) the inverse of V in its exact form; (iii) the integral and rational powers of V and V^{-1}.

Thus we have, from the elementary rule on the inversion of a product of several matrices,

$$\begin{aligned}
\sigma^2 V^{-1} = (I+\alpha_1 U'+\ldots+\alpha_k U'^k)(I+\beta_1 U'+\ldots+\beta_h U'^h)^{-1} \\
\times (I+\beta_1 U+\ldots+\beta_h U^h)^{-1}(I+\alpha_1 U+\ldots+\alpha_k U^k),
\end{aligned} \tag{33}$$

which is no more difficult to evaluate exactly than the matrix product in (31), which is the autocovariance matrix itself.

It is apparent from (31) and (33) that although V is a Laurent matrix, V^{-1} is not, a property already noted by Whittle.

* See also Wise (1955, Chapter III) for a further analysis of this problem.

The autoregressive process

To illustrate the application of the above results to problems of importance which have not been solved in the literature even in special cases, the inverse of the autocovariance matrix of the autoregressive process generated by the relation

$$x_t + \alpha_1 x_{t-1} + \ldots + \alpha_k x_{t-k} = \epsilon_t \tag{34}$$

will now be derived for the semi-infinite process.

From this result the inverses of the autocovariance matrices of N successive observations from the process (34) are then given in full for $k = 1$, $k = 2$ and $k = 3$.

From (33) we deduce for the special case (34) the result

$$\sigma^2 V^{-1} = (I + \alpha_1 U' + \ldots + \alpha_k U'^k)(I + \alpha_1 U + \ldots + \alpha_k U^k). \tag{35}$$

Taking $k = 1$, 2 and 3 respectively, we deduce at once from (35) the matrices

(i) $k = 1$:

$$\sigma^2 V^{-1} = \begin{bmatrix} 1 & \alpha_1 & 0 & \ldots & 0 & 0 \\ \alpha_1 & 1+\alpha_1^2 & \alpha_1 & \ldots & 0 & 0 \\ 0 & \alpha_1 & 1+\alpha_1^2 & \ldots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \ldots & 1+\alpha_1^2 & \alpha_1 \\ 0 & 0 & 0 & \ldots & \alpha_1 & 1 \end{bmatrix} \tag{36}$$

an $N \times N$ matrix.

(ii) $k = 2$:

$$\sigma^2 V^{-1} = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & 0 & \ldots & 0 & 0 \\ \alpha_1 & 1+\alpha_1^2 & \alpha_1+\alpha_1\alpha_2 & \alpha_2 & \ldots & 0 & 0 \\ \alpha_2 & \alpha_1+\alpha_1\alpha_2 & 1+\alpha_1^2+\alpha_2^2 & \alpha_1+\alpha_1\alpha_2 & \ldots & 0 & 0 \\ 0 & \alpha_2 & \alpha_1+\alpha_1\alpha_2 & 1+\alpha_1^2+\alpha_2^2 & \ldots & 0 & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1+\alpha_1^2 & \alpha_1 \\ 0 & 0 & 0 & 0 & \ldots & \alpha_1 & 1 \end{bmatrix} \tag{37}$$

an $N \times N$ matrix.

(iii) $k = 3$:

$$\sigma^2 V^{-1} = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & \alpha_3 & \ldots & 0 & 0 \\ \alpha_1 & 1+\alpha_1^2 & \alpha_1+\alpha_1\alpha_2 & \alpha_2+\alpha_1\alpha_3 & \ldots & 0 & 0 \\ \alpha_2 & \alpha_1+\alpha_1\alpha_2 & 1+\alpha_1^2+\alpha_2^2 & \alpha_1+\alpha_1\alpha_2+\alpha_2\alpha_3 & \ldots & 0 & 0 \\ \alpha_3 & \alpha_2+\alpha_1\alpha_3 & \alpha_1+\alpha_1\alpha_2+\alpha_2\alpha_3 & 1+\alpha_1^2+\alpha_2^2+\alpha_3^2 & \ldots & 0 & 0 \\ \vdots & & & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1+\alpha_1^2 & \alpha_1 \\ 0 & 0 & 0 & 0 & \ldots & \alpha_1 & 1 \end{bmatrix}. \tag{38}$$

The result (36) confirms that obtained for $k = 1$ by Cochrane & Orcutt (1949, equation (6·14), p. 57). However, the results (37) and (38) for $k = 2$ and $k = 3$ respectively do not appear to have been given previously in the literature.* So far as the author is aware, (35) is new, and the inverses of the autocovariance matrices for $k = 4, 5, 6, \ldots$, etc., can be written down immediately from it, rendering 'approximate' methods of inversion unnecessary.

* Note, however, Champernowne (1948, p. 206, equation 3·5), which gives the inverse implicitly for the case of the autoregressive process.

REFERENCES

ANDERSON, R. L. (1942). Distribution of the serial correlation coefficient. *Ann. Math. Statist.* **13**, 1.

CHAMPERNOWNE, D. G. (1948). Sampling theory applied to autoregressive schemes. *J. R. Statist. Soc.* **10**, 204.

COCHRANE, D. & ORCUTT, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *J. Amer. Statist. Ass.* **44**, 32.

DOOB, J. L. (1953). *Stochastic Processes.* New York: John Wiley and Sons.

JORDAN, C. (1947). *Calculus of Finite Differences.* New York: Chelsea Publishing Co.

KHINTCHINE, A. (1934). Korrelationstheorie der stationären stochastischen Prozesse. *Math. Ann.* **109**, 604.

WATSON, G. S. (1951). Serial correlation in regression analysis. Unpublished Ph.D. thesis and Inst. of Stat. Mimeo. Series, North Carolina State College.

WHITTLE, P. (1951). *Hypothesis Testing in Time-Series Analysis.* Uppsala: Almquist and Wiksell.

WHITTLE, P. (1952). Some results in time-series analysis. *Skand. ActuarTidskr.* **35**, 200.

WIENER, N. (1930). Generalized harmonic analysis. *Acta-Math.* **55**, 117.

WISE, J. (1955). Ph.D. Thesis (unpublished), University of London.

WOLD, H. (1953). *Demand Analysis.* New York: John Wiley and Sons.

WOLD, H. (1954). *A Study in the Analysis of Stationary Time Series.* Uppsala: Almquist and Wiksell.

SAMPLING PROPERTIES OF LOCAL STATISTICS IN STATIONARY STOCHASTIC SERIES

By G. H. JOWETT

University of Sheffield

1. Introduction

The sampling moments of large-sample statistics calculated from stationary time series usually require for their evaluation the summation of long series of cumulants or products of cumulants of the series, or for normal series the summation of long series of autocorrelations or products of autocorrelations (cf. Bartlett, 1946). These series usually converge only as the cumulants in question tend to zero, which is often in practice so slow a process as to make their evaluation a matter of considerable difficulty. There is, however, a class of statistics which may be called *local* in that they are dependent on short-term comparisons of terms in the series, and for these the summations are of such a kind as to converge as certain finite differences of the cumulants tend to zero; in practice this often happens much more quickly than the tending to zero of the cumulants themselves, and the corresponding formulae are accordingly much easier to evaluate in practice.

A simple illustration of this difference is provided by a comparison between the simplest statistic which depends on local comparisons and the simplest statistic which does not. It is advantageous to give this illustration before attempting a general discussion, so as to give a simple outline of the argument and the ideas involved.

Suppose $x_1, x_2, \ldots, x_{2n-1}, x_{2n}$ to be a set of $2n$ evenly spaced observations from a stationary time series with mean μ, variance σ^2, and autocorrelation function ρ_s. The sampling variance of the statistic

$$U \equiv n^{-1}[(x_1 - x_2) + (x_3 - x_4) + \ldots + (x_{2n-1} - x_{2n})] \tag{1}$$

is determined as follows:

$$\operatorname{var} U \equiv n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \operatorname{cov}\left[(x_{2i-1} - x_{2i}), (x_{2j-1} - x_{2j})\right] \tag{2}$$

$$= n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma^2(-\rho_{2(i-j)-1} + 2\rho_{2(i-j)} - \rho_{2(i-j)+1}) \tag{3}$$

$$= n^{-1} \sum_{s=-(n-1)}^{+(n-1)} \left(1 - \frac{|s|}{n}\right)(-\Delta'' \sigma^2 \rho_{2s}), \tag{4}$$

where Δ'' is the second central difference operator in Comrie's notation, and refers to the suffix of ρ. The need for this statistic and its sampling variance might arise, for example, in the estimation of the treatment effect in a systematic experiment having the pattern

$$T\ C\ T\ C\ \ldots\ T\ C \quad (T \equiv \text{treatment}; C \equiv \text{control}), \tag{5}$$

or, alternatively, in the statistical control of the difference between the means of two interpenetrating systematic samples (Jowett, 1955 b).

On the other hand, the sampling variance of the simple arithmetic mean

$$T \equiv (n^{-1})(x_1 + x_3 + \ldots + x_{2n-1}) \tag{6}$$

is easily shown to be given by the formula

$$\operatorname{var} T = \frac{1}{n} \sum_{s=-(n-1)}^{+(n-1)} \left(1 - \frac{|s|}{n}\right) \sigma^2 \rho_s. \tag{7}$$

In many applications, as $|s|$ increases the autocorrelation function tends to become 'locally linear', i.e. approximately linear over intervals of width equal, say, to twice the interval between successive terms of the series; as $|s|$ increases further it gradually tends to zero. Thus $\Delta''\rho_{2s}$ may attain a sufficient degree of smallness to be neglected in computing the approximate value of the sum in (4) at a much earlier stage in the increase of $|s|$ than that for which ρ_{2s} can be neglected in computing the approximate value of the sum in (7).

Now experience and *a priori* considerations (cf. Jowett, 1953) suggest that for many series observed in practice the graph of the serial variation function

$$\delta_s = E[\tfrac{1}{2}(x_t - x_{t+s})^2] \tag{8}$$

is zero for $s = 0$, rises steeply as s increases with gradually decreasing gradient and curvature, and ultimately flattens out when s becomes large, rather like the graph of the function

$$y = A(1 - e^{-ks})$$

for positive values of the constants A and k.

Since

$$\delta_s = \sigma^2(1 - \rho_s), \tag{9}$$

it follows that

$$-\Delta''\sigma^2\rho_s = \Delta''\delta_s, \tag{10}$$

and hence that the graph of $\sigma^2\rho_s$ will have the same shape but inverted. The serial variation function has the advantage that for short lags it is effectively independent of long-term variation in the series and requires no reference to the series mean, qualities which make it eminently suitable for use in the study of local variation. If for $s \geqslant s_0(\epsilon)$ the serial variation function has become approximately linear to such an extent that $\Delta''\delta_{2s}$ is less than some predetermined small quantity ϵ, we deduce from (4) that

$$\operatorname{var} U = \frac{1}{n} \sum_{s=-(s_0-1)}^{+(s_0-1)} \left(1 - \frac{|s|}{n}\right) \Delta''\delta_{2s} + R, \tag{11}$$

where

$$|R| < \frac{2\epsilon}{n} \sum_{s=s_0}^{(n-1)} 1 - \frac{|s|}{n} < \epsilon. \tag{12}$$

Hence the error in the estimate of var U committed by truncating the summation at $\pm (s_0 - 1)$ is at worst of the same order of magnitude as the departure from local linearity of δ_s over the rest of the range, which is usually the same as that of the first term omitted from the summation, since $|\Delta''\delta_{2s}|$ will usually tend to decrease monotonically for $s > s_0$.

Since we are usually obliged to infer the stochastic properties from one or at most a very few realizations of the series itself, it will often be true that var U is much easier to estimate than var T; the need to estimate the form of ρ_s for large lags, often a matter of considerable difficulty, is not present, and the formula (4) may be computed approximately (provided, of course, that $\epsilon = o(n^{-1})$ and $s_0 \ll n$) using serial variation statistics with relatively short lags, which are usually readily available.

It will be observed that U is a function of the set of differences $x_1 - x_2, x_3 - x_4, \ldots$, which might be described as *local comparisons*, since they are constructed from adjacent terms of

II

the series. It is this which gives rise to the differencing in (4), which makes the formula independent of σ^2, and hence expressible entirely in terms of δ_s, and which enables var U to be estimated in many circumstances where it would be impossible to estimate var T. A corresponding property has been observed in other statistics, such as the mean of a systematic sample (cf. Jowett, 1952) or a trend-reduced regression coefficient (Jowett, 1955a) which are functions of local comparisons alone, and suggests that the sampling properties of statistics which are built up entirely from local comparisons are essentially determined by the local variational properties (e.g. δ_s for small s) of the series themselves, and are effectively independent of the long-term variational properties of the series which are so difficult to measure in practice. The rest of this paper will be concerned with establishing this principle, and with illustrating it by some specific formulae.

2. Sampling moments of local statistics in stationary normal series

We shall assume a reference space S with position vector $\mathbf{t}$, and shall use the symbol τ to denote a displacement vector, e.g. $\mathbf{t}_1 - \mathbf{t}_2$. This is a generalization of the usual time axis in one dimension. A set of stochastic variables $x_1(\mathbf{t}), x_2(\mathbf{t}), \ldots$ (which may be regarded as components of a vector random function) with respective means $\mu_1, \mu_2, \ldots$ and variances $\sigma_1^2, \sigma_2^2, \ldots$, is defined at the points of S; in practice the definition is often at a lattice of points only, but for generality we shall assume definition at all points of S. These variables will be taken to form stationary series, i.e. their probability parameters to be invariant under translation in S.

We shall be concerned with *seminvariant local linear functions* (s.l.l.f.'s) of these variables defined as follows:

$$L_{\alpha i} \equiv \int_s x_i(\mathbf{t})\, dL_\alpha(\mathbf{t}), \quad \int_s dL_\alpha(\mathbf{t}) = 0. \tag{13}$$

The integration is to be understood in the Stieltjes sense, so that

$$L_{\alpha i} \equiv \sum_r l_{\alpha r}^* x_i(\mathbf{t}_r) + \int_s x_i(\mathbf{t})\, l_\alpha^{**}(\mathbf{t})\, d\mathbf{t}, \tag{14}$$

where $l_{\alpha r}^*$ is zero except at a finite number of points of S, and $l^{**}(\mathbf{t})$ is integrable in the ordinary Riemann sense. These linear functions of $x(\mathbf{t})$ we called *seminvariant* because they are unchanged by the addition of any constant to $x(\mathbf{t})$; to justify the adjective *local*, $l_{\alpha r}^*$ and $l_\alpha^{**}(\mathbf{t})$ will be taken as zero outside a minimal sphere in S of diameter a_α bounded above by some fixed value a. These s.l.l.f.'s are generalizations of the simple local comparisons implied in § 1.

It will be assumed that the series are multivariate normal. This assumption may, however, be dispensed with when we are concerned only with second moments of linear statistics, since the question of normality does not then arise. The assumption of normality implies that

$$Ex_1(\mathbf{t}_1)\,x_2(\mathbf{t}_2)\ldots x_s(\mathbf{t}_s) = \begin{cases} \mu_1\mu_2\ldots\mu_s + \text{other terms each having at least one of } \mu_1\mu_2\ldots\mu_s \\ \quad \text{as a factor} \quad (s \text{ odd}), \\[6pt] \mu_1\mu_2\ldots\mu_3 + \sum\limits_{\alpha\beta\ldots\nu} \mathrm{cov}_{x_\alpha x_\beta}(\mathbf{t}_\alpha - \mathbf{t}_\beta)\,\mathrm{cov}_{x_\gamma x_\delta}(\mathbf{t}_\gamma - \mathbf{t}_\delta)\ldots \mathrm{cov}_{x_\mu x_\nu}(\mathbf{t}_\mu - \mathbf{t}_\nu) \\ \quad + \text{other terms each having at least one of } \mu_1\mu_2\ldots\mu_s \\ \quad \text{as a factor} \quad (s \text{ even}), \end{cases} \tag{15}$$

where $\alpha\beta \ldots \mu\nu$ is any permutation of $l \ldots s$ such that

$$\alpha < \beta, \gamma < \delta, \ldots, \mu < \nu \quad \text{and} \quad \alpha < \gamma \ldots < \mu. \tag{16}$$

If we prefer to work with lag variation parameters defined by

$$\delta_{x_i x_j}(t - t') = E \tfrac{1}{2}(x_i(t) - x_j(t'))^2, \tag{17}$$

we have

$$\delta_{x_i x_j}(t - t') = \tfrac{1}{2}(\mu_i - \mu_j)^2 + \tfrac{1}{2}(\sigma_i^2 + \sigma_j^2) - \operatorname{cov}_{x_i x_j}(t - t') \tag{18}$$

as the relation which permits us to substitute δ for cov. In analysis concerned with s.l.l.f.'s, the seminvariance usually results in the elimination of the terms involving means and variances. For example,

$$\operatorname{cov}(L_{\alpha i}, L_{\beta j}) = E(L_{\alpha i} L_{\beta j}) \quad (\text{since } E(L_{\alpha i}) = 0)$$

$$= \iint_s x_i(t) x_j(t') \, dL_\alpha(t) \, dL_\beta(t')$$

$$= \iint_s \operatorname{cov}_{x_i x_j}(t - t') \, dL_\alpha(t) \, dL_\beta(t')$$

$$= \iint_s -\delta_{x_i x_j}(t - t') \, dL_\alpha(t) \, dL_\beta(t'). \tag{19}$$

The double integration is interpretable in the obvious way in terms of summations and Riemann integrations.

In addition to the assumption of normality, we shall make a further important assumption about the nature of the lag variation functions (l.v.f.'s), an assumption which is very often reasonable in practice (cf. Jowett, 1953). We shall assume that the l.v.f.'s tend to become increasingly linear as $|\tau|$ increases; or, more precisely, that for arbitrary (small) $\epsilon > 0$, and for arbitrary c, there is a value $h_0(\epsilon, c)$ such that for $h > h_0$ and $h < |\tau| < h + c$, we can find a constant ψ_{ij} and a constant vector χ_{ij} such that

$$\delta_{x_i x_j}(\tau) = \psi_{ij} + \chi_{ij} . \tau + \epsilon_{ij}(\tau), \tag{20}$$

where

$$|\epsilon_{ij}(\tau)| < \epsilon. \tag{21}$$

Two s.l.l.f.'s will be described as *far apart* if their associated minimal spheres touch the outside of a sphere of diameter greater than some predetermined distance h_0. If we take h_0 as dependent on ϵ as described above, and take c as twice the maximum local diameter a, it follows that if L_α, L_β are far apart,

$$\operatorname{cov}(L_{\alpha i}, L_{\beta j}) = \iint -\delta_{x_i x_j}(t - t') \, dL_\alpha(t) \, dL_\beta(t')$$

$$= -\iint [\psi_{ij} + \chi_{ij} . (t - t') + \epsilon_{ij}(t - t')] \, dL_\alpha(t) \, dL_\beta(t')$$

$$= -\iint \epsilon_{ij}(t - t') \, dL_\alpha(t) \, dL_\beta(t'), \tag{22}$$

and hence that

$$|\operatorname{cov}(L_{\alpha i}, L_{\beta j})| \leqslant \epsilon \int_s |dL_\alpha(t)| \int_s |dL_\beta(t)|$$

$$= O(\epsilon). \tag{23}$$

On the other hand, if L_α and L_β are not far apart, $\operatorname{cov}(L_{\alpha i}, L_{\beta j})$ is a function of $\delta_{x_i x_j}(\tau)$ over the range

$$0 \leqslant |\tau| < h_0 + 2a, \tag{24}$$

i.e. in what may be called a $(h_0 + 2a)$ neighbourhood of $\tau = 0$. If h_0 and a are sufficiently small (in practice, h_0 is often comparable with a in magnitude), the behaviour of $\delta_{x_i x_j}(\tau)$ in such a neighbourhood may be interpreted as a local variational property of the series.

This property of being either dependent on the l.v.f. in a neighbourhood of $\tau = 0$ or of magnitude $O(\epsilon)$ is also true of the cross-moment of several s.l.l.f.'s. For the cross-moment of s of these functions, which have zero expectations, we have

$$E(L_{1i_1}, L_{2i_2} \ldots L_{si_s}) = \int\int \ldots \int E[x_{i_1}(\mathbf{t}) \, x_{i_2}(\mathbf{t}') \ldots x_{i_s}(\mathbf{t}^{(s-1)})] \, dL_1(\mathbf{t}) \, dL_2(\mathbf{t}') \ldots dL_s(\mathbf{t}^{(s-1)})$$

$$= \begin{cases} 0 & \text{if } s \text{ is odd,} \\ \int\int \ldots \int (-1)^{\frac{1}{2}s} \sum_{\alpha\beta\ldots\mu\nu} [\delta_{(\alpha\beta)}(\mathbf{t} - \mathbf{t}') \, \delta_{(\gamma\delta)}(\mathbf{t}'' - \mathbf{t}''') \ldots \delta_{(\mu\nu)}(\mathbf{t}^{(s-2)} - \mathbf{t}^{(s-1)})] \\ \qquad \times dL_\alpha(\mathbf{t}) \ldots dL_\nu(\mathbf{t}^{(s-1)}) \text{ if } s \text{ is even,} \end{cases} \tag{25}$$

(where the suffix $(\alpha\beta)$ means $x_{i_\alpha} x_{i_\beta}$), since all the terms in the expectation of the product in the square bracket which have one or more of $\mu_1 \ldots \mu_s$ as factor are annihilated in the integration by the seminvariant property of the corresponding L's.

The summation in (25) is taken over the set of permutations of $1, 2, \ldots, s$ defined in (16). Hence if s is even,

$$E(L_{1i_1} \ldots L_{si_s}) = \sum_{\alpha\beta\ldots\nu} \text{cov}(L_{\alpha i_\alpha}, L_{\beta i_\beta}) \, \text{cov}(L_{\gamma i_\gamma}, L_{\delta i_\delta}) \ldots \text{cov}(L_{\mu i_\mu}, L_{\nu i_\nu}). \tag{26}$$

Since the right-hand side of (26) is a function of covariances, it follows that $E(L_{1i_1} L_{2i_2} \ldots L_{si_s})$ is either dependent on the l.v.f. in a neighbourhood of $\tau = 0$ or of magnitude at most of order ϵ.

If the series $x_{i_1}(\mathbf{t}), \ldots$ are made to coincide in sets, and the s.l.l.f.'s $L_{1i_1}, L_{2i_2}, \ldots$ to coincide in sets included in them, the result just established takes the following form:

Theorem. The expectation of any product of powers of s.l.l.f.'s of any set of concomitant stationary stochastic variables for which the s.v.f.'s and the l.v.f.'s have a tendency towards linearity with increasing $|\tau|$ (as defined by (20) and (21) and the preceding paragraph) is a sum of terms which are either of magnitude at most of order ϵ or involve only values of the s.v.f.'s and l.v.f.'s in a $(h_0 + 2a)$ neighbourhood of $\tau = 0$.

It follows that the moments and cumulants of any statistic which is a product of powers of s.l.l.f.'s have the same property, since these are expressible as linear functions of products of powers; and hence that the moments and cumulants of any statistic which can be expressed as a linear function of products of powers has the same property. The theorem is thus seen to be very general in character, for most if not all statistics which measure local properties and are of interest in practice may be expressed either exactly or approximately in this form. The theorem implies, broadly speaking, that the variational properties of local statistics depend essentially on local variational properties of the parent series.

3. Asymptotic moments of large-sample local statistics

Suppose that the statistics $L_{1i_1} \ldots L_{si_s}$ (s even and divisible by an integer p) fall into r sets of p, namely, $\mathcal{M}_1 \ldots \mathcal{M}_r$. Suppose, moreover, that the product of the statistics falling into the set $\mathcal{M}_i$ is denoted by M_i. Thus

$$M_1 = L_{1i_1} L_{2i_2} \ldots L_{pi_p}, \quad M_2 = L_{p+1, i_{p+1}} L_{p+2, i_{p+2}} \ldots L_{2p, i_{2p}}, \quad \ldots \quad M_r = L_{r-1 \cdot p+1, i_{r-1 \cdot p+1}} \ldots L_{s, i_s}. \tag{27}$$

It is easily shown that

$$E(M_1 - E(M_1))(M_2 - E(M_2)) \dots (M_r - E(M_r))$$
$$= \sum_{\alpha', \dots, \mu'} \operatorname{cov}(L_{\alpha t_\alpha}, L_{\beta t_\beta}) \operatorname{cov}(L_{\gamma t_\gamma}, L_{\delta t_\delta}) \dots \operatorname{cov}(L_{\mu t_\mu}, L_{\nu t_\nu}), \qquad (28)$$

where $\alpha\beta \dots \mu\nu$ is a permutation of the integer $1 \dots \kappa$ subject to the conditions (16) and also to the condition that no pair of suffixes associated with a covariance included in the formula may come from the same set M.

Two sets will be said to be *far apart* if the members of one set are far apart from the members of the other. If any set is far apart from all the others, the product moment (28) will be $O(\epsilon^p)$, since every member of the set has to be associated in a covariance with some member of another set. For some of the terms in (28) not to be $O(\epsilon^p)$, for each set there must be at least one other set from which it is not far apart; if this is so, terms which are not $O(\epsilon^p)$ arise, but only when members from such neighbouring sets occur together in the covariance brackets.

Many statistics which are of practical interest are means, or functions of means, of powers or products of powers of s.l.l.f.'s of the stochastic variables over a sample region of the space S or at a set of sample points, usually evenly spaced. Suppose that we are given sets $M_1, M_2, \dots, M_n$, where n is large (for example, of the same order of magnitude as the number of sample points), and where the sets are *evenly spread* over the sample region in such a way that every subregion of volume $O(V/n)$ (i.e. $O(1)$) contains sets to the number $O(1)$. Then if

$$U = (n^{-1})(M_1 + M_2 + \dots + M_n), \qquad (29)$$

we have

$$E(U - E(U))^r = (n^{-r}) E \sum_{\alpha_1, \dots, \alpha_r} (M_{\alpha_1} - E(M_{\alpha_1})) \dots (M_{\alpha_r} - E(M_{\alpha_r})). \qquad (30)$$

where $\alpha_1 \dots \alpha_r$ is a permutation of r of the integers $1 \dots n$.

If we assume $h \ll n$, the number of terms in (30) which are not $O(\epsilon^p)$, i.e. which can be formed from products of covariances of s.l.l.f.'s which are not far apart, is $O(n^{\frac{1}{2}r})$, since the number of ways of choosing pairs of sets from $M_1, M_2 \dots M_n$ which are not far apart is of this order. Hence

$$E(U - E(U))^r = O(n^{-\frac{1}{2}r}) + O(\epsilon^p), \qquad (31)$$

the second term on the right-hand side of (31) being justified by the fact that the summation in (30) has only n^r terms altogether, and is divided by n^r.

If $\epsilon^p = o(n^{-\frac{1}{2}r})$, the moment will be dominated by those terms which involve only values of the l.v.f. in a $(h + 2a)$ neighbourhood of $\tau = 0$, and will be of magnitude $O(n^{-\frac{1}{2}r})$.

This argument may be generalized in a fairly obvious way to show that cross-moments of order r about the mean for statistics having the same form as U also have this property and this order of magnitude. Thus, since many statistics of useful potential application have sampling errors which may be expressed, to a large-sample approximation at least, as linear functions of products of sampling errors of statistics having this form, it may be shown that the lower moments of these also depend essentially on values of the l.v.f. in a restricted neighbourhood of $\tau = 0$.

4. EXAMPLES

Example 1. *Variance of a trend-reduced covariance*

Suppose we have a sample stretch $x_1 y_1, x_2 y_2, \dots, x_n y_n$ from the variation of two concomitant series, x, y. The covariance of the difference between successive terms, which is the simplest case of a trend-reduced covariance (cf. Jowett, 1955a), is defined by the equation

$$U = (n-1)^{-1}[(x_1 - x_2)(y_1 - y_2) + \dots + (x_{n-1} - x_n)(y_{n-1} - y_n)]. \qquad (32)$$

Then

$$\operatorname{var} U = (n-1)^{-2} \sum_{i,j=1}^{n-1} \operatorname{cov}\left[(x_i - x_{i+1})(y_i - y_{i+1}), (x_j - x_{j+1})(y_j - y_{j+1})\right]$$

$$= (n-1)^{-2} \sum_{i,j=1}^{n-1} \operatorname{cov}\left[(x_i - x_{i+1}), (x_j - x_{j+1})\right] \operatorname{cov}\left[(y_i - y_{i+1}), (y_j - y_{j+1})\right]$$
$$+ \operatorname{cov}\left[(x_i - x_{i+1}), (y_j - y_{j+1})\right] \operatorname{cov}\left[(y_i - y_{i+1}), (x_j - x_{j+1})\right]$$

$$= (n-1)^{-2} \sum_{i,j=1}^{n-1} \Delta''\delta_{xx}(i-j)\Delta''\delta_{yy}(i-j) + [\Delta''_{xy}(i-j)]^2$$

$$= (n-1)^{-1} \sum_{s=-(n-2)}^{+(n-2)} \left(1 - \frac{|s|}{n-1}\right)\{\Delta''\delta_{xx}(s)\Delta''\delta_{yy}(s) + [\Delta''_{xy}(s)]^2\}. \tag{33}$$

For independent series, $\delta_{xy}(s)$ is constant for all s, so that

$$\Delta''\delta_{xy}(s) = 0. \tag{34}$$

The dominant terms in the summation are those for small $|s|$; as $|s|$ increases, and the δ's straighten out, the remaining terms will rapidly tend to zero.

Example 2. *The covariance of two serial variation statistics*

The serial variation statistics of lags s_1 and s_2 ($s_1 \leqslant s_2$) calculated from a stretch of series $x_1 \ldots x_n$ are defined as follows:

$$d(s_1) = (n - s_1)^{-1} \sum_{i=1}^{n-s_1-1} \tfrac{1}{2}(x_i - x_{i+s_1})^2,$$

$$d(s_2) = (n - s_2)^{-1} \sum_{i=1}^{n-s_2-1} \tfrac{1}{2}(x_i - x_{i+s_2})^2. \tag{35}$$

Then

$$\operatorname{cov}(d(s_1), d(s_2))$$

$$= (n-s_1)^{-1}(n-s_2)^{-1} \sum_{i=1}^{n-s_1-1}\sum_{j=1}^{n-s_2-1} \operatorname{cov}\left(\tfrac{1}{2}(x_i - x_{i+s_1})^2, \tfrac{1}{2}(x_j - x_{j+s_2})^2\right)$$

$$= (n-s_1)^{-1}(n-s_2)^{-1} \sum_{i,j} 2[\operatorname{cov}\left(\tfrac{1}{2}(x_i - x_{i+s_1}), \tfrac{1}{2}(x_j - x_{j+s_2})\right)]^2$$

$$= (n-s_1)^{-1}(n-s_2)^{-1} \sum_{\alpha=-(n-s_1-1)}^{n-s_1-1} (n - c_\alpha) \cdot \tfrac{1}{2}[-\delta(\alpha) - \delta(\alpha + s_2 - s_1) + \delta(\alpha - s_1) + \delta(\alpha + s_2)]^2$$

$$= n^{-1} \sum_{\alpha=-(n-s_1-1)}^{n-s_1-1} \frac{n(n-c_\alpha)}{(n-s_1)(n-s_2)} \cdot \tfrac{1}{2}[-\delta(\alpha) - \delta(\alpha + s_2 - s_1) + \delta(\alpha - s_1) + \delta(\alpha + s_2)]^2, \tag{36}$$

where

$$c_\alpha = \begin{cases} s_1 + |\alpha|, & s_2 - s_1 \leqslant \alpha \leqslant n - s_1 - 1, \\ s_2, & 0 \leqslant \alpha < s_2 - s_1, \\ s_2 + |\alpha|, & -(n - s_2 - 1) \leqslant \alpha \leqslant 0. \end{cases} \tag{37}$$

The dominant terms in the summation in (36) are again those for small $|s|$. As $|s|$ increases, and $\delta(s)$ straightens out, the remaining terms, being proportional to squares of a kind of second difference, again tend rapidly to zero.

Example 3. *Spatial systematic sampling*

The systematic sampling will be of the variable $x(u, v)$ which is defined over the rectangular region $0 \leqslant u \leqslant r_1$, $0 \leqslant v \leqslant r_2$, samples being taken at the points $(i - \tfrac{1}{2}, j - \tfrac{1}{2})$ for integral i, j such that $i = 1, \ldots, r_1, j = 1, \ldots, r_2$.

Let

$$L_{ij} = x(i - \tfrac{1}{2}, j - \tfrac{1}{2}) - \int_{u=i-1}^{i} \int_{v=j-1}^{j} x(u, v)\, du\, dv. \tag{38}$$

Then

$$\text{cov}(L_{ij}, L_{i'j'}) = -\delta(i - i', j - j') - \int_{\theta=i-i'-1}^{i-i'+1}\int_{\phi=j-j'-1}^{j-j'+1}(1 - |\theta - i + i'|)(1 - |\phi - j + j'|)\,\delta(\theta, \phi)\,d\theta\,d\phi$$

$$+ \int_{i'-1}^{i'}\int_{j'-1}^{j'}\delta(i - \tfrac{1}{2} - u, j - \tfrac{1}{2} - v)\,du\,dv + \int_{i-1}^{i}\int_{j-1}^{j}\delta(u - i' - \tfrac{1}{2}, v - j' - \tfrac{1}{2})\,du\,dv. \quad (39)$$

This covariance is itself a s.l.l.f. of $\delta(\theta, \phi)$ over the region

$$i - i' - 1 \leq \theta \leq i - i' + 1, \quad j - j' - 1 \leq \phi \leq j - j' + 1. \quad (40)$$

and will be negligible if the points (i, j), (i', j') are far enough apart.

The sampling variance of the mean U of a systematic sample of n ($= r_1 r_2$) about the mean of the rectangular region is therefore given by

$$\text{var}\, U = \text{var}\left(n^{-1}\sum_{i=1}^{r_1}\sum_{j=1}^{r_2}L_{ij}\right) \quad (41)$$

$$= n^{-2}\sum_{iji'j'=1}^{n}\text{cov}(L_{ij}, L_{i'j'}) = (n^{-2})\sum_{iji'j'}c(i - i', j - j'). \quad \text{say}. \quad (42)$$

In evaluating the summation in (42), we take the terms in order of increasing distance between (i, j), (i', j'). Hence

$$\text{var}\, U = \frac{1}{n}\left\{c(0, 0) + \left[\frac{2n}{(r_1 - 1)r_2}c(1, 0) + \frac{2n}{r_1(r_2 - 1)}c(0, 1)\right] + \frac{4n}{(r_1 - 1)(r_2 - 1)}c(1, 1)\ldots\right\}. \quad (43)$$

As the distance increases and $\delta(\theta, \phi)$ becomes locally linear, the terms will become negligible.

In Ex. 1, 2 and 3 we are dealing with second moments, corresponding to $r = 2$ in (30). Hence we verify that they are all of order $n^{-\frac{1}{2}r} \equiv n^{-1}$, provided that local linearity is attained rapidly enough. The next example is different, involving a greater value of r.

Example 4. Fourth moment of the statistic $U \equiv (n^{-1})\{(x_1 - x_2) + (x_3 - x_4) + \ldots + (x_{2n-1} - x_{2n})\}$

The statistic U is identical with that defined in (1). In the expansion of U^4, the expectation of a typical term is given by

$$E(x_{2i-1} - x_{2i})(x_{2i'-1} - x_{2i'})(x_{2i''-1} - x_{2i''})(x_{2i'''-1} - x_{2i'''})$$
$$= \Delta''\delta(2\overline{i - i'})\Delta''\delta(2\overline{i'' - i'''}) + \Delta''\delta(2\overline{i - i''})\Delta''\delta(2\overline{i' - i'''}) + \Delta''\delta(2\overline{i - i'''})\Delta''\delta(2\overline{i' - i''})$$
$$\quad (44)$$

$$\equiv U_{ii'i''i'''}, \quad \text{say}.$$

The value of this depends only on the configuration of i, i', i'', i'''. If these are placed in increasing order of magnitude, we may denote the central interval by q, and the other two by p and r, where $p \leq r$. The number of terms giving the same values of p, q and r, and hence the same expectations, is equal to $(n - p - q - r)$ multiplied by the factor a_{pqr} given below:

Range of p, q, r	a_{pqr}
$0 < p < r$; $q > 0$	$2.4!$
$0 < p = r$; $q > 0$	$4!$
$0 = p < r$; $q > 0$	$2.4!/2!$
$0 = p = r$; $q > 0$	$4!/(2!)^2$
$0 < p < r$; $q = 0$	$2.4!/2!$
$0 < p = r$; $q = 0$	$4!/2!$
$0 = p < r$; $q = 0$	$2.4!/3!$
$0 = p = q = r$	1

If we write the total of these terms in each case as $(n-p-q-r)\,B_{pqr}$, we obtain the following result:

$$E(U^4) = n^{-4} \sum_{i,\,i',\,i'',\,i'''=1}^{n} u_{ii'i''i'''}$$

$$= n^{-4} \sum_{\substack{p,\,q,\,r \\ (p\leqslant r,\,p+q+r\leqslant n-1)}} (n-p-q-r)\,B_{pqr}$$

$$= n^{-4} \sum_{r=0}^{n-1} \sum_{p=0}^{\min(r,\,n-r-1)} \sum_{q=0}^{n-p-r} (n-p-q-r)\,B_{pqr}. \tag{45}$$

If for $s \geqslant s_0$, $\delta(2s)$ differs from linearity over the interval $(2s-1,\,2s+1)$ by less than some small quantity ϵ, we have

$$|\Delta''\delta(2s)| < 4\epsilon. \tag{46}$$

Since
$$B_{pqr} \propto \Delta''\delta(2p)\,\Delta''\delta(2r) + \Delta''\delta(\overline{2p+q})\,\Delta''\delta(2q+r) + \Delta''\delta(\overline{2p+q+r})\,\Delta''(2q), \tag{47}$$

B_{pqr} will be of magnitude $O(1)$ if both p and r are less than s_0, of magnitude $O(\epsilon)$ if just one of p, q, r is less than s_0, and of magnitude $O(\epsilon^2)$ if none is less than s_0.

The summation in (45) may be separated into parts having different orders of magnitude. If $s_0 \ll n$,

$$E(U^4) = n^{-4} \sum_{r=0}^{s_0-1} \sum_{p=0}^{r} \sum_{q=0}^{n-r-p-1} (n-p-q-r)\,B_{pqr}$$

$$+ n^{-4} \sum_{r=s_0}^{n-1} \sum_{p=0}^{s_0-1} \sum_{q=0}^{n-r-p-1} (n-p-q-r)\,B_{pqr}$$

$$+ n^{-4} \sum_{r=s_0}^{n-1} \sum_{p=s_0}^{\min(r,\,n-r-1)} \sum_{q=0}^{\min(s_0-1,\,n-r-p-1)} (n-p-q-r)\,B_{pqr}$$

$$+ n^{-4} \sum_{r=s_0}^{n-1} \sum_{p=s_0}^{\min(r,\,n-r-1)} \sum_{q=\min(s_0-1,\,n-r-p-1)}^{n-r-p-1} (n-p-q-r)\,B_{pqr}. \tag{48}$$

The four terms on the right-hand side of (48) are of magnitudes respectively

$$O(n^{-2}\epsilon^0), \quad O(n^{-1}\epsilon^1), \quad O(n^{-1}\epsilon^1) \quad \text{and} \quad O(n^0\epsilon^2).$$

Hence if
$$\epsilon = o(n^{-1}) \tag{49}$$

the right-hand side of (48) will be dominated by the first term. To order n^{-2}, this may be put in a more readily computable form:

$$E(U^4) \sim n^{-2} \sum_{r=0}^{s_0-1} \sum_{p=0}^{r} \tfrac{1}{2} a_{p1r} \Delta''\delta(2p)\,\Delta''\delta(2r), \tag{50}$$

since
$$\sum_{q=1}^{n-r-p-1} (n-p-q-r) = \tfrac{1}{2}(n-r-p)(n-r-p-1) \sim \tfrac{1}{2}n^2, \tag{51}$$

and $a_{pqr} = a_{p1r}$ for $q > 0$. To this order, we may neglect the term in a_{p0r}.

This example has been chosen because of its comparative simplicity, to illustrate the way in which the summations in these problems have to be arranged in order to take the important terms first. It is a fourth moment, corresponding to $r = 4$ in (30), and we have verified that it is of order $n^{-\frac{1}{2}r} = n^{-2}$, provided that linearity in $\delta(s)$ is attained rapidly enough. It might be required in allowing for the effect of terms of order n^{-2} in the test of

significance of the treatment effect in (5) (cf. Small, 1954), or in obtaining control chart limits correct to order n^{-1} for the difference between means of interpenetrating systematic samples from, say, a conveyor belt.

REFERENCES

BARTLETT, M. S. (1946). Some aspects of the time-correlation problem in regard to tests of significance. *J.R. Statist. Soc.* **98**, 536.

JOWETT, G. H. (1952). The accuracy of systematic sampling from conveyor belts. *Appl. Statist.* **1**, 50.

JOWETT, G. H. (1953). The comparison of means for industrial time series. Paper given to Royal Statistical Society conference.

JOWETT, G. H. (1955*a*). Least squares regression analysis for trend-reduced time series. *J.R. Statist. Soc.* B (in the Press).

JOWETT, G. H. (1955*b*). The comparison of means of industrial time series. *Appl. Statist.* **4** (in the Press).

SMALL, V. J. (1954). M.Sc. Thesis for the University of Sheffield.

MODELS FOR TWO-DIMENSIONAL STATIONARY STOCHASTIC PROCESSES

By V. HEINE

Applied Mathematics Laboratory, Department of Scientific and Industrial Research, Wellington, N.Z.[*]

To aid the analysis of two-dimensional stationary processes, three different models are considered, derived from the second-order stochastic partial differential equation. Their correlation functions are calculated, for fitting data in the form of a correlogram to the model. The corresponding Green functions describe the physical nature of the models, and in particular distinguish between space-like and time-like axes.

1. INTRODUCTION

The occurrence of stationary stochastic processes in two dimensions is well known. In analysing such data, it is at least useful, and occasionally of theoretical importance, to fit the correlogram to a particular plausible model and to estimate certain parameters in the model accordingly. However, owing to mathematical complexities, only the correlation functions generated by the very simplest models have been investigated in the past, particularly by Whittle (1954), and it would seem desirable to extend the range of available models.

We shall consider the general second-order linear stochastic partial differential equation

$$\left(a\frac{\partial^2}{\partial x^2} + 2h\frac{\partial^2}{\partial x \partial y} + b\frac{\partial^2}{\partial y^2} + 2g\frac{\partial}{\partial x} + 2f\frac{\partial}{\partial y} + c \right) \xi(x,y) = \epsilon(x,y), \tag{1.1}$$

where ξ and ϵ are the variate and the random impulses effecting it respectively, both with means zero. This leads to three types of model, corresponding to parabolic, elliptic and hyperbolic forms, whose correlation and Green functions we shall derive.

In two-dimensional processes, one is necessarily concerned with the difference between time-like and space-like axes. In a time series, the variate can only be influenced by past events; and accordingly, a time-like axis is one, such that a random impulse can only produce an effect in one direction along it. Besides time itself, examples are the distance down a steep hillside, and the distance in the direction of a strong wind scattering seeds. Along a space-like axis, the variate depends on events in both directions. We discuss these features for each model in terms of the Green function, which represents the effect at one point, of a random impulse at another point. In fact, the whole nature of the process is best visualized by considering the Green function, which thus shows what kind of physical processes the models may represent.

2. PRELIMINARY THEORY

The results of §§ 2 and 3 are analogous to those relating to one-dimensional time series, and we present them here but briefly in the particular form required later (cf. Bartlett (1946) and Daniell (1946)).

Consider the stochastic partial differential equation

$$L\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right) \xi(x,y) = \epsilon(x,y). \tag{2.1}$$

[*] Now at Sidney Sussex College, Cambridge.

Assuming the validity of inverting the order of differentiation and integration, the solution may be written formally

$$\xi(x,y) = \int_{-\infty}^{\infty} \int_{x}^{\infty} G(x-u, y-v)\, \epsilon(u,v)\, du\, dv, \qquad (2\cdot2)$$

where the Green function $G(x,y)$ satisfies

$$L\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) G(x,y) = \delta(x)\,\delta(y), \qquad (2\cdot3)$$

using the Dirac delta function (cf. Van der Pol & Bremner, pp. 75, 315). The physical interpretation of (2·2) and (2·3) is that $G(x-u, y-v)$ represents the effect at the point (x,y) of a unit impulse at the point (u,v).

If the $\epsilon(u,v)$ are entirely uncorrelated random impulses, we have for the covariance functions

$$\mathrm{cov}_\epsilon(x,y) = \text{expectation of } [\epsilon(u,v)\,\epsilon(u-x, v-y)]$$
$$= \sigma^2 \delta(x)\,\delta(y), \qquad (2\cdot4)$$

and

$$\mathrm{cov}_\xi(x,y) = \sigma^2 \int_{x}^{\infty} \int_{-\infty}^{\infty} G(u,v)\, G(u-x, v-y)\, du\, dv \qquad (2\cdot5)$$

by using (2·2) and (2·4). In what follows, it will be convenient to calculate the function

$$R(x,y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(u,v)\, G(u-x, v-y)\, du\, dv, \qquad (2\cdot6)$$

which can then be normalized right at the end, to give the correlation function of ξ in the form

$$\rho(x,y) = R(x,y)/R(0,0). \qquad (2\cdot7)$$

Further, G and R must tend to zero at infinity, and R must be finite everywhere.

3. Use of the Laplace transform

The manipulative complexities, encountered in calculating the Green and correlation functions from (2·3), (2·6) and (2·7), are handled using the two-sided Laplace transform. We follow completely the notation and dictionary of formulae in Van der Pol & Bremner (1950), reference to whom will be made simply by the page numbers. Thus if $f(p,q)$ is the transform of $h(x,y)$ we write (pp. 18, 334–6)

$$h(x,y) \doteqdot f(p,q),$$

where

$$f(p,q) = pq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-px-qv} h(x,y)\, dx\, dy.$$

If $g(p,q)$ be written for the transform of $G(x,y)$, (2·3) gives (pp. 48, 345, 384)

$$L(p,q)\, g(p,q) = pq,$$

i.e.

$$G(x,y) \doteqdot g(p,q) = \frac{pq}{L(p,q)}. \qquad (3\cdot1)$$

The transform of (2·6) may be written down using (3·1) and the composition product formula (pp. 39, 382); thus

$$R(x,y) \doteqdot \frac{g(p,q)\, g(-p,-q)}{pq} = \frac{pq}{L(p,q)\, L(-p,-q)}. \qquad (3\cdot2)$$

4. STANDARD FORMS

By the simple substitutions $x' = -x, kx, \pm y, x\cos\theta + y\sin\theta$ and in the hyperbolic case by $x' = x + c_1 y, y' = x + c_2 y$; the general form (1·1) can always be reduced to one of the following standard forms:

parabolic
$$\left(\frac{\partial}{\partial x} + \alpha\right)^2 - \gamma^2\left(\frac{\partial}{\partial y} \pm \beta\right); \tag{4·1}$$

elliptic
$$\left(\frac{\partial}{\partial x} - \alpha\right)^2 + \frac{\partial^2}{\partial y^2} \pm \gamma^2; \tag{4·2}$$

hyperbolic
$$\left(\frac{\partial}{\partial x} + \alpha\right)\left(\frac{\partial}{\partial y} + \beta\right) \pm \gamma^2; \tag{4·3}$$

degenerate
$$\gamma\frac{\partial^2}{\partial y^2} + \alpha\frac{\partial}{\partial y} \pm \beta; \tag{4·4}$$

where α, β, γ are all real and positive, or zero.

5. THE PARABOLIC FORM

Consider
$$L\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = \left(\frac{\partial}{\partial x} + \alpha\right)^2 - \gamma^2\left(\frac{\partial}{\partial y} + \beta\right). \tag{5·1}$$

From (3·1)
$$G(x, y) = \frac{pq}{(p+\alpha)^2 - \gamma^2(q+\beta)}, \tag{5·2}$$

hence
$$e^{\alpha x + \beta y} G(x, y) = \frac{pq}{p^2 - \gamma^2 q} \quad \text{(p. 374)}$$

$$= \frac{-p}{\gamma^2}\exp\left(\frac{p^2 y}{\gamma^2}\right) U(y) \quad \text{(p. 26)},$$

hence
$$G(x, y) = \frac{-1}{2\gamma\sqrt{(\pi y)}}\exp\left(-\alpha x - \beta y - \frac{x^2\gamma^2}{4y}\right) U(y) \quad \text{(p. 21)}, \tag{5·3}$$

where
$$U(y) = 1 \quad (y \geqslant 0); \tag{5·4}$$
$$= 0 \quad (y < 0).$$

The case of the lower sign in (4·1) is covered by considering β as negative in (5·1), which makes G in (5·3) non-vanishing at infinity, and is thus inadmissible.

Equation (5·3) shows that an impulse at the origin only has an effect at positive values of y. Thus the y-axis is a time-like axis, whereas the x-axis is space-like. The Green function has the shape of a Gaussian error curve in the x-direction, with ever-increasing variance and decreasing amplitude as y increases. Hence the y-axis may well represent distance downhill, downstream or along the direction of a wind in any kind of diffusion phenomenon. Otherwise it may represent time, during which something spreads out along a line, for instance, the descendants of a plant.

From (3·2),

$$\gamma^4 R(x, y) = \frac{pq}{\left(\dfrac{\gamma^2\beta - \alpha^2 - p^2}{\gamma^2}\right)^2 - \left(q - \dfrac{2\alpha p}{\gamma^2}\right)^2}$$

$$= \tfrac{1}{2}\exp\left(\frac{2\alpha p}{\gamma^2}y\right)\frac{\gamma^2 p}{(\gamma^2\beta - \alpha^2 - p^2)}\exp\left[\frac{(p^2 + \alpha^2 - \gamma^2\beta)}{\gamma^2}|y|\right] \quad \text{(p. 27)}. \tag{5·5}$$

Without loss of generality consider the two quadrants with $y > 0$. Thus rearranging,

$$R(x,y) = \frac{1}{2\gamma^2} e^{-\beta q} \left| \frac{p}{p+\alpha} (p+\alpha) \exp\left[\frac{(p+\alpha)^2 y}{\gamma^2} \right] \right| \left| \frac{p}{(\gamma^2\beta - x^2) - p^2} \right| \frac{1}{p}. \tag{5.6}$$

Now

$$\frac{p}{p+\alpha} (p+\alpha) \exp\left[\frac{(p+\alpha)^2 y}{\gamma^2} \right] \cdot \frac{1}{2\gamma\sqrt{(\pi y)}} \exp\left(-\frac{x^2\gamma^2}{4y} \right) \quad \text{(p. 24)}, \tag{5.7}$$

and

$$\frac{p}{(\gamma^2\beta - x^2) - p^2} \cdot \frac{1}{2\sqrt{(\gamma^2\beta - x^2)}} \exp\left(-\sqrt{(\gamma^2\beta - \alpha^2)} |x| \right) \quad \text{(p. 27)}. \tag{5.8}$$

We can now use the composition product rule (p. 39), also (5.7) and (5.8), in (5.6):

$$R(x,y) = (\text{constant}) \int_{\sqrt{y}}^{} \int_{-x}^{} \exp\left(x\tau - \frac{\gamma^2\tau^2}{4y} - \sqrt{(\gamma^2\beta - x^2)} |x - \tau| \right) d\tau. \tag{5.9}$$

The integral in (5.9) may be simplified by breaking it up into two integrals over the ranges $\tau < x$ and $\tau > x$ respectively, and then using linear substitutions

$$t = \frac{\gamma}{2\sqrt{y}} \left(\tau + \frac{2xy}{\gamma^2} \right) \pm \sqrt{\left| y\left(\beta - \frac{x^2}{\gamma^2} \right) \right|}.$$

Using (2.7) to insert the appropriate normalizing constant, we finally have

$$\rho(x,y) = \rho(-x,-y), \quad \text{where} \quad y > 0,$$
$$= \frac{e^{-2AB}}{\sqrt{\pi}} \int_{-\infty}^{A-B} e^{-t^2} dt$$
$$+ \frac{e^{+2AB}}{\sqrt{\pi}} \int_{A+B}^{\infty} e^{-t^2} dt, \tag{5.10}$$

where

$$A = \frac{\gamma}{2\sqrt{y}} \left(x + \frac{2xy}{\gamma^2} \right), \quad B = \sqrt{\left| y\left(\beta - \frac{x^2}{\gamma^2} \right) \right|}.$$

with

$$0 \leqslant \alpha^2 < \beta\gamma^2 \quad \text{[see (5.13)]}.$$

As emphasized by Van der Pol & Bremner (1950, pp. 19, 27), to ensure the validity of the above argument, it is necessary to show that (5.2)–(5.9) possess a common non-vanishing region of convergence. This requirement eliminates several other solutions of (5.2) and (5.5) which would otherwise appear to be formally possible, and also imposes some sufficient conditions on α, β and γ. By considering (5.1), (2.3) and (2.6) directly, without the use of the Laplace transform, these conditions may be shown to be also necessary.

Thus (5.3) requires (pp. 26, 374)

$$\text{Re}\left[-\beta + \left(\frac{p+\alpha}{\gamma} \right)^2 \right] < \text{Re}\, q < \infty. \tag{5.11}$$

(5.5) in addition requires (p. 378)

$$-\infty < \text{Re}\, q < -\text{Re}\left[-\beta + \left(\frac{p-\alpha}{\gamma} \right)^2 \right]. \tag{5.12}$$

For these regions to overlap, we must have the left-hand expression in (5.11) less than the right-hand expression in (5.12), which leads to

$$0 \leqslant \alpha^2 < \beta\gamma^2. \tag{5.13}$$

It may be shown that if and only if (5.13) holds, do (5.2) to (5.9) have a non-vanishing common region of convergence.

6. Elliptic forms

The special form

$$L\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} - \gamma^2 \tag{6.1}$$

has been treated by Whittle (1954), who used it in discussing the yields of orange trees in a square array. For the sake of completeness, we repeat

$$G(x, y) = \frac{1}{2\pi} K_0(\gamma r) \doteq \frac{pq}{p^2 + q^2 - \gamma^2} \quad \text{(pp. 357, 407)}. \tag{6.2}$$

$R(x, y)$ is obtained by differentiating (6.2) with respect to γ (p. 373). Then from (2.7),

$$\rho(x, y) = (\gamma r) K_1(\gamma r). \tag{6.3}$$

Here and elsewhere, $r = \sqrt{(x^2 + y^2)}$, and the K's are the modified Bessel functions of the second kind. Both the x- and y-axes are space-like, and for (6.1) the Green and correlation functions decrease monotonically equally in all directions. A simple change of scale in the x- or y-direction introduces unequal degrees of correlation along these axes. This model would be applicable to a wide variety of circumstances, such as field trials on flat land.

A form involving a further parameter is

$$L\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = \left(\frac{\partial}{\partial x} - \alpha\right)^2 + \frac{\partial^2}{\partial y^2} - \gamma^2. \tag{6.4}$$

From (3.1), (6.2) and pp. 357, 339,

$$G(x, y) = \frac{e^{\alpha x}}{2\pi} K_0(\gamma r) \doteq \frac{pq}{(p - \alpha)^2 + q^2 - \gamma^2}. \tag{6.5}$$

This Green function is the same as (6.2), except that the term $\exp(\alpha x)$ increases the influence an impulse has in the positive x-direction at the expense of the negative x-direction, and thus may represent the effect of a wind or a slight slope of the ground. From (3.2),

$$R(x, y) \doteq \frac{1}{4\alpha p} \left[\frac{pq}{(p - \alpha)^2 + q^2 - \gamma^2} - \frac{pq}{(p + \alpha)^2 + q^2 - \gamma^2} \right].$$

Using (6.5), the integration rule (p. 51) and (2.7)

$$\rho(x, y) = \frac{\sqrt{(\gamma^2 - \alpha^2)}}{\sin^{-1}(\alpha/\gamma)} \int_x^\infty \sinh(\alpha \tau) K_0(\gamma \sqrt{(\tau^2 + y^2)}) \, d\tau, \tag{6.6}$$

where the normalizing constant follows from a formula in Watson (1944) on p. 388. Convergence of (6.6) requires $\alpha < \gamma$.

All other elliptic forms are inadmissible. Consider

$$L\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = \left(\frac{\partial}{\partial x} - \alpha\right)^2 + \frac{\partial^2}{\partial y^2} + \gamma^2. \tag{6.7}$$

If $\alpha = \gamma = 0$, the Green function is $\log r$ which does not vanish at infinity. If $\alpha = 0$, $\gamma \neq 0$, the Green function is $\frac{1}{4} Y_0(\gamma r)$ (p. 357), and the integral in (2.6) does not converge. If $\alpha \neq 0$, $\gamma \neq 0$, the Green function $\frac{1}{4} \exp(\alpha x) Y_0(\gamma r)$ does not vanish at infinity.

7. HYPERBOLIC FORMS

Consider

$$L\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = \left(\frac{\partial}{\partial x} + \alpha\right)\left(\frac{\partial}{\partial y} + \beta\right) + \gamma^2. \tag{7.1}$$

From (3.1) and p. 346, and using the function U defined in (5.4),

$$G(x, y) = e^{-\alpha x - \beta y} J_0(2\gamma \sqrt{(xy)}) U(x) U(y)$$

$$\doteqdot \frac{pq}{(p+\alpha)(q+\beta)+\gamma^2} \tag{7.2}$$

is the only Green function that vanishes at infinity. Its form shows that any impulse can only have an influence in the positive x- and y-directions, which therefore represent two time-like axes.
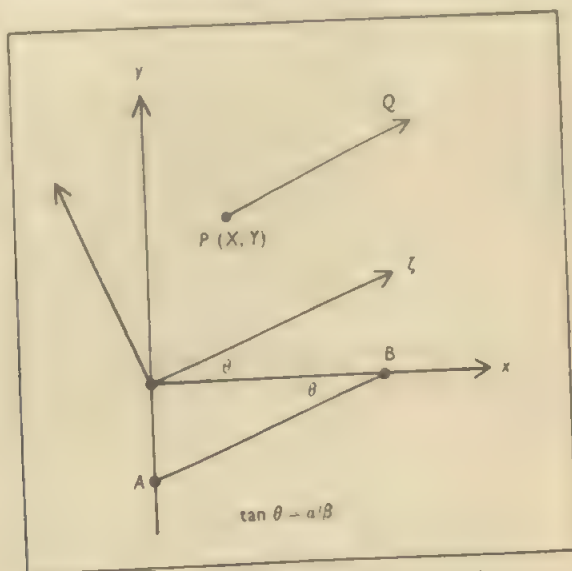
Fig. 1. Path of integration for (7.7).

From (3.2),

$$R(x, y) \doteqdot \frac{-1}{2(\beta p + \alpha q)}\left[\frac{pq}{(p+\alpha)(q+\beta)+\gamma^2} - \frac{pq}{(p-\alpha)(q-\beta)+\gamma^2}\right], \tag{7.3}$$

with region of convergence $|\operatorname{Re} p| < \alpha$, $|\operatorname{Re} q| < \beta$. If we put

$$f(x, y) = e^{-\alpha x - \beta y} J_0(2\gamma \sqrt{(xy)}) U(x) U(y) - e^{\alpha x + \beta y} J_0(2\gamma \sqrt{(xy)}) U(-x) U(-y)$$

$$\doteqdot \left[\frac{pq}{(p+\alpha)(q+\beta)+\gamma^2} - \frac{pq}{(p-\alpha)(q-\beta)+\gamma^2}\right], \tag{7.4}$$

then (7.3) becomes

$$2\left(\beta\frac{\partial}{\partial x} + \alpha\frac{\partial}{\partial y}\right) R(x, y) = -f(x, y). \tag{7.5}$$

Hence $R(X, Y)$ is obtained by integrating $f(x, y)$ of (7.4) in the ζ direction, i.e. along PQ (Fig. 1), making an angle $\tan^{-1}(\alpha/\beta)$ with the x-axis:

$$R(X, Y) = \frac{1}{2\sqrt{(\alpha^2 + \beta^2)}}\int_{P(X, Y)}^{\infty} f(x, y)\, d\zeta. \tag{7.6}$$

In the first quadrant, $f(x, y)$ is equal to the first term in (7·4), and

$$\rho(X, Y) = \frac{2\alpha\beta}{\surd(\alpha^2 + \beta^2)} \sqrt{\left(1 + \frac{\gamma^2}{\alpha\beta}\right)} \int_{P(X, Y)}^{\infty} e^{-\imath x \cdot \beta v} J_0(2\gamma \surd(xy)) \, d\zeta. \tag{7·7}$$

where the normalizing constant is obtained from a formula on p. 384 in Watson (1944). In the fourth quadrant, $f(x, y)$ is zero, so that ρ is constant along any line AB (Fig. 1), and is equal to $\rho(B)$ as given by (7·7). In the third and fourth quadrants, ρ is obtained from the relation $\rho(X, Y) = \rho(-X, -Y)$. It is necessary for the convergence of (7·7) that $\alpha > 0$ and $\beta > 0$. In evaluating (7·7) numerically, it would be helpful to employ a change of scale to make α and β each $1/\surd 2$.

The form

$$L\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = \left(\frac{\hat{e}}{\partial x} + \alpha\right) \left(\frac{\partial}{\partial y} + \beta\right) \tag{7·8}$$

may be considered as a special case of (7·1) with $\gamma = 0$, but it is easier to start again from first principles. From (3·1) and p. 386,

$$G(x, y) = e^{-\alpha x - \beta v} U(x) U(y) \doteqdot \frac{pq}{(p + \alpha)(q + \beta)}. \tag{7·9}$$

From (3·2), (2·7) and p. 386,

$$\rho(x, y) = \exp\left(-\alpha |x| - \beta |y|\right) \tag{7·10}$$

over all four quadrants.

Considering the form

$$L\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = \left(\frac{\partial}{\partial x} + \alpha\right) \left(\frac{\partial}{\partial y} + \beta\right) - \gamma^2, \tag{7·11}$$

the above analysis (7·1) to (7·7) applies, except that J_0 must now be replaced by I_0, and the sign of γ^2 changed. In this case it is necessary that $\gamma^2 < \alpha\beta$ for convergence.

8. Degenerate cases

Degenerate forms

$$L_1\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) = \gamma \frac{\partial^2}{\partial y^2} + \alpha \frac{\partial}{\partial y} \pm \beta \tag{8·1}$$

are strictly speaking inadmissible. For from (3·2),

$$R(x, y) = \delta(x) R_1(y), \tag{8·2}$$

where $R_1(y)$ is some function of y; and thus $R(x, y)$ cannot be normalized in the sense of (2·7), because of the delta function.

However, a case may well arise in practice, where the correlogram is approximately zero everywhere except along one direction, say the y-axis. It is therefore necessary to consider forms arbitrarily close to (8·1); or what is the same thing, to consider (8·1) as the result of some limiting process.

The important fact is that the result depends on the particular limiting process used. How this is possible may be seen as follows. Consider the rather restricted form

$$L\left(\eta \frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) \quad \text{with limit (8·1) as } \eta \to 0. \tag{8·3}$$

The correlation function may always be written in the form

$$\rho(x, y) = \rho(0, y) N(x/\eta, y), \tag{8·4}$$

where $\qquad$ $N(x/\eta, y) = 1$ for $x = 0$,

$\qquad$ $= \text{small}$ for $x \gg \eta$, since $\rho \to 0$ as $x/\eta \to \infty$.

In the limit, the form (8·1) has the correlation function

$$\rho(x, y) = \rho_1(y) N(x),$$ (8·5)

where the null function

$$N(x) = 1 \quad (x = 0);$$
$$= 0 \quad (x \neq 0).$$

As there are many different forms satisfying (8·3), so there are different correlation functions (8·4) and (8·5). We give the following example.

Limiting form: $\qquad\qquad \dfrac{\partial}{\partial y} + \beta.$

I form: $\qquad\qquad \left(\eta \dfrac{\partial}{\partial x} + 1\right)\left(\dfrac{\partial}{\partial y} + \beta\right).$

Limiting correlation function, from (7·10):

$$\rho(x, y) = e^{-\beta|y|} N(x).$$

II form: $\qquad\qquad \left(-\eta^2 \dfrac{\partial^2}{\partial x^2}\right) + \left(\dfrac{\partial}{\partial y} + \beta\right).$

Limiting correlation function, from (5·1):

$$\rho(x, y) = \frac{2N(x)}{\sqrt{\pi}} \int_{\sqrt{(\beta|y|)}}^{\infty} e^{-t^2} dt.$$

Corresponding results for other more general forms of the type (8·1) may be easily obtained.

9. SUMMARY OF ALLOWED MODELS

The allowed standard forms (cf. §4) are as follows:

Parabolic; x-axis space-like, y-axis time-like:

$$\left(\frac{\partial}{\partial x} + \alpha\right)^2 - \gamma^2\left(\frac{\partial}{\partial y} + \beta\right),$$
$$0 \leqslant \alpha^2 < \beta\gamma^2;$$

Green function (5·3), correlation function (5·10):

Elliptic; x- and y-axes space-like:

$$\left(\frac{\partial}{\partial x} - \alpha\right) + \frac{\partial^2}{\partial y^2} - \gamma^2,$$
$$0 \leqslant \alpha < \gamma;$$

Green function (6·5), correlation function (6·6):

Hyperbolic; x- and y-axes time-like:

$$\left(\frac{\partial}{\partial x} + \alpha\right)\left(\frac{\partial}{\partial y} + \beta\right) + \gamma^2,$$
$$\alpha > 0, \quad \beta > 0, \quad \gamma^2 \geqslant 0;$$

Green function (7·2), correlation function (7·7): also

$$\left(\frac{\partial}{\partial x}+\alpha\right)\left(\frac{\partial}{\partial y}+\beta\right)-\gamma^2,$$

$$\alpha>0, \quad \beta>0, \quad 0\leqslant\gamma^2<\alpha\beta;$$

for Green and correlation functions, see above (7·11).

The problem was suggested by Dr P. Whittle, to whom I am also indebted for helpful discussion.

REFERENCES

BARTLETT, M. S. (1946). *Stochastic Processes.* Mimeographed, North Carolina lecture notes.

DANIELL, P. J. (1946). Contribution to discussion on stochastic processes. *J. R. Statist. Soc.* B, 8, 88.

VAN DER POL, B. & BREMNER, H. (1950). *Operational Calculus based on the Two-sided Laplace Transform.* Cambridge University Press.

WATSON, G. N. (1944). *The Theory of Bessel Functions,* 2nd ed. Cambridge University Press.

WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika,* 41, 434.

SOME PROBLEMS IN THE THEORY OF PROVISIONING AND OF DAMS

By J. GANI

Australian National University, Canberra, A.C.T.

The paper begins with a review of two problems in the theory of provisioning with a discrete stock considered by Pitt (1946), and a short account of Moran's work (1954) in the theory of finite dams. It is pointed out that these provide two different methods of attack, each appropriate to certain conditions, on problems in the probability theory of a general storage function $S(t)$ defined at time t by

$$S(t) = I(t) - D(t) - F(t),$$

where $I(t)$, $D(t)$, $F(t)$ are respectively an input, output and overflow function. This storage function is identified with the stock deficit in provisioning theory, or the dam content in dam theory, so that any problem and its solution in the one theory has an exact analogue in the other.

The paper continues with the use of Pitt's results of the theory of provisioning in two analogous cases of the infinite discrete dam. This is followed by the application of Moran's methods of the theory of dams in some analogous problems of provisioning with a discrete finite stock, exact solutions are obtained for the discrete and continuous cases of a particular problem in which ordering and replacement times coincide.

The paper closes with an exact solution of the general storage problem in the case of a finite continuous storage function $S(t)$, fed by a discrete input function of Poisson type, with a continuous output which has a steady rate when $S(t) \neq 0$, and is zero if $S(t) = 0$, and an overflow function such that $S(t)$ never exceeds a prescribed value.

I. PROBLEMS IN THE THEORY OF PROVISIONING

In his work on the theory of provisioning, Pitt (1946) is concerned with the derivation of stationary probability distributions for a discrete stock function $s(t)$. This stock function represents the number of components in stock at a time t, and is allowed to take integral values ranging between K and $-\infty$, where $K = s(0)$ is the initial stock, and negative values indicate that components may be borrowed if they are not in stock. The function is defined for all values of time t (t continuous) by the equation

$$s(t) = K + r(t) - c(t), \tag{1}$$

where $c(t)$ is the consumption function, the number of components consumed up to time t, a random increasing step function taking positive integral values, and $r(t)$ is the replacement function, the number of components delivered and added to the stock up to time t, a function depending on $c(t)$ and always smaller than or equal to it.

The probability distribution associated with the consumption function $c(t)$ is the Poisson with parameter a, so that the probabilities that in a small interval of time δt there be one and no components consumed respectively are

$$\left. \begin{array}{l} \Pr\{c(t+\delta t) - c(t) = 1\} = a\,\delta t, \\ \Pr\{c(t+\delta t) - c(t) = 0\} = 1 - a\,\delta t, \end{array} \right\} \tag{2}$$

and increases in $c(t)$ in non-overlapping intervals are independent. Two types of replacement functions are considered:

Type 1.
$$r_1(t) = [c(t-T)\,M^{-1}]\,M, \tag{3}$$

where we define $[x]$ as the integral part of x; here, orders for a constant number of components M are sent out at irregular time intervals $t - T$, when $c(t-T)$ is an integral multiple of M, and are delivered and added to the stock at time t, after a positive time-lag T;

Type 2. $$r_2(t) = c\{[(t-T)aM^{-1}]\,Ma^{-1}\};\qquad(4)$$

here, orders for an irregular number of components equal to the consumption in the previous time interval Ma^{-1} are sent out at regular times $k\,Ma^{-1}$ ($k = 1, 2, \ldots$), and are delivered and added to stock at times $k\,Ma^{-1} + T$ after a positive time-lag T.

For both these types of replacement function, the stationary probability distributions of $s(t)$ are derived by arguments which will be followed exactly in §IV of the paper. Using these probability distributions, Pitt concludes with a comparison of the means of both the positive and negative values of the stock function $s(t)$ when the replacement functions are of types 1 and 2.

II. Problems in the theory of dams

The problem discussed by Moran (1954), that of obtaining the stationary probability distribution of the amount of water in a dam, appears at first sight to be of a different nature. A finite dam of capacity K units, whose content at the times $t = 0, 1, 2, \ldots$, after water has been released according to a prescribed rule, is $Z(t)$, where $Z(t)$ will be found to lie in the range $(0, K - M)$, is subject to the following conditions:

(1) a discrete input $X(t)$ which flows into the dam during the interval of time $(t, t+1)$, where the series $\{X(t)\}$ is serially independent, and the probability that the input be i units ($i = 0, 1, 2, \ldots$) is p_i;

(2) an overflow rule such that in any time interval $(t, t+1)$ there is no overflow from the dam if $Z(t) + X(t) \leqslant K$, and there is an overflow $Z(t) + X(t) - K$ if $Z(t) + X(t) > K$;

(3) a release rule such that at time $t+1$, M units are released from the dam if its total content $Z(t) + X(t) \geqslant M$, and the total content $Z(t) + X(t)$ is released if this is less than M.

Then, provided $K - M \geqslant M$, a system of equations is derived relating the probabilities $P_0, P_1, \ldots, P_{K-M}$ that $Z(t)$ be equal to $0, 1, \ldots, K - M$ at time t, with the probabilities $P_0^{(1)}, P_1^{(1)}, \ldots, P_{K-M}^{(1)}$ that $Z(t+1)$ be equal to $0, 1, \ldots, K - M$ at time $t+1$. These are:

$$\left.\begin{aligned}
P_0^{(1)} &= P_0(p_0 + p_1 + \ldots + p_M) + P_1(p_0 + \ldots + p_{M-1}) + \ldots + P_M p_0,\\
P_1^{(1)} &= P_0 p_{M+1} \qquad\quad + P_1 p_M \qquad\qquad + \ldots + P_{M+1} p_0,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
P_{K-M-1}^{(1)} &= P_0 p_{K-1} \qquad\quad + P_1 p_{K-2} \cdots\cdots + \ldots + P_{K-M} p_{M-1},\\
P_{K-M}^{(1)} &= P_0 q_K \qquad\qquad + P_1 q_{K-1} \qquad\quad + \ldots + P_{K-M} q_M,
\end{aligned}\right\}\qquad(5)$$

where, for convenience, we have written $q_j = \sum_{i=j}^{\infty} p_i$. Written in matrix form, these equations are $\mathbf{P}^{(1)} = \mathbf{p}\mathbf{P}$, where $\mathbf{P}$, $\mathbf{P}^{(1)}$ are column vectors with elements P_i, $P_i^{(1)}$ respectively, and $\mathbf{p}$ is the matrix of coefficients.

The stationary distribution $\{P_i\}$ is required, and this is obtained by writing $\mathbf{P}^{(1)} = \mathbf{P}$ in the set of equations (5) and solving $\mathbf{P} = \mathbf{p}\mathbf{P}$ together with the additional condition $\sum_{i=0}^{K-M} P_i = 1$. It is pointed out that although no solution in an explicit form may exist in general, the values of the stationary probabilities P_i can always be evaluated numerically for any known matrix of coefficients $\mathbf{p}$, and these are in fact computed for a special case. An extension of this discrete theory also allows equations for the continuous stationary distribution of $U(t) = Z(t) + X(t)$ to be written when $Z(t)$, $X(t)$ are continuous and $U(t)$ lies in the range $(0, \infty)$.

III. Provisioning and Dam Problems as General Storage Problems

We now proceed to identify these problems of the stock function $s(t)$ of provisioning, and of the dam content $Z(t)$ in dam theory as particular problems of a general storage function $S(t)$. This may be discrete or continuous, and defined for continuous time t or at fixed times such as $t = 0, 1, 2, \ldots$, according to the conditions of the problem, by the equation

$$S(t) = I(t) - D(t) - F(t), \tag{6}$$

where $I(t)$, $D(t)$, $F(t)$ are positive increasing functions, discrete or continuous. $I(t)$ is a random input function which feeds the storage function, $D(t)$ is an output function which depletes the storage function and may depend on $I(t)$ or be otherwise defined, and $F(t)$ is an overflow function which operates to restrict the range of $S(t)$ below a prescribed maximum value.

To show that the problems of Pitt's stock function $s(t)$ of provisioning can be regarded as problems of a particular storage function $S(t)$, we consider the previously defined stock function (1),

$$s(t) = K + r(t) - c(t),$$

which lies in the range $-\infty < s(t) \leqslant K$. We note that this equation holds at any time t (t continuous), and that $c(t)$, the random discrete consumption function, and $r(t)$, the discrete replacement function $(r(t) \leqslant c(t))$, are defined as in §1. We can here equally well consider the stock deficit,

$$K - s(t) = c(t) - r(t), \tag{7}$$

which ranges from zero when the stock is at its maximum K, to ∞ as the stock decreases through 0 to $-\infty$. Naturally this is possible only when the stock has a maximum value K which it cannot exceed. Comparing this with a storage function

$$S(t) = I(t) - D(t), \tag{8}$$

for which $F(t) = 0$ for all t, it is clear that a discrete storage function $S(t)$ defined in the range $(0, \infty)$ for all time t (t continuous), is identifiable with the deficit $K - s(t)$. Similarly, $I(t)$, a discrete random input function, is identifiable with the discrete random consumption function $c(t)$, and $D(t)$, a discrete output function $(D(t) \leqslant I(t))$, with a discrete replacement function $r(t)$. $F(t)$ will be zero at all times, since the range of the stock deficit is $(0, \infty)$, and no overflow is required to restrict this range below any prescribed value.

In the specific problems considered by Pitt, where the consumption function $c(t)$ was of Poisson type with parameter a, and the replacement functions of the types previously defined in (3), (4), as $r_1(t)$ and $r_2(t)$, the general storage equation would be such that

$$S(t) = K - s(t), \quad I(t) = c(t), \quad D(t) = r_1(t) \quad \text{or} \quad r_2(t), \quad F(t) = 0.$$

However, generally, for any stock function $s(t)$, whether continuous or discrete, defined for continuous time t or at discrete intervals of time, providing only that the maximum stock K is finite, it is always possible to frame the general storage equation (6) for the stock deficit. The finite maximum stock K results in the storage function $S(t)$ having a range with lower bound 0; its upper bound may be infinite, as in the case just considered, or finite as we shall see in §V. $I(t)$, $D(t)$ and $F(t)$ will be appropriately chosen to fit the conditions of the problem.

Similarly, we proceed to show that Moran's discrete dam content $Z(t)$ with range $(0, K - M)$ can be regarded as a particular storage function. In this case, however, the function $Z(t)$

is defined only at the times $t = 0, 1, 2, ...$, and we shall also define the storage function at these times. In our storage equation (6), we may identify $S(t)$ with $Z(t) - Z(0)$, the increase in dam content after a time t; if, without loss in generality, we assume that $Z(0) = 0$, that is, the dam is empty at time $t = 0$, we have that $S(t)$ the storage function is equal to the dam content $Z(t)$. The random discrete input function $I(t)$ would be the sum of the discrete random inputs $X(0), X(1), ..., X(t-1)$ in the time intervals $(0, 1), (1, 2), ..., (t-1, t)$, so that

$$I(t) = \sum_{\tau=1}^{t} X(\tau-1).$$

The discrete output function $D(t)$ would be the sum of discrete outputs released at times $\tau = 1, 2, ..., t$; let these be written $d(1), d(2), ..., d(t)$, then we have that

$$d(\tau) = M \quad \text{if} \quad Z(\tau-1) + X(\tau-1) \geqslant M,$$

or

$$d(\tau) = Z(\tau-1) + X(\tau-1) \quad \text{if this is less than } M,$$

so that the output function is

$$D(t) = \sum_{\tau=1}^{t} d(\tau).$$

Finally, the overflow function $F(t)$ would be the sum of the discrete overflows in the time intervals $(0, 1), ... (t-1, t)$; let these be written $f(0), ..., f(t-1)$, then we have that

$$f(\tau-1) = 0 \quad \text{if} \quad Z(\tau-1) + X(\tau-1) \leqslant K,$$

or

$$f(\tau-1) = Z(\tau-1) + X(\tau-1) - K \quad \text{if this is greater than zero,}$$

so that the overflow function is

$$F(t) = \sum_{\tau=1}^{t} f(\tau-1).$$

The equation for $Z(t)$ at times $t = 0, 1, 2, ...$ could be written

$$Z(t) = \sum_{\tau=1}^{t} X(\tau-1) - \sum_{\tau=1}^{t} d(\tau) - \sum_{\tau=1}^{t} f(\tau-1), \tag{9}$$

a particular case of the storage function $S(t)$. For a continuous dam content, a similar equation could also be written. More generally, for a $Z(t)$ continuous or discrete, defined at any time t (t continuous) or at discrete times, with a finite or infinite range, it is possible to frame a general storage equation.

In neither the case of problems of provisioning nor of dam problems will the storage equation itself necessarily provide a method of solution enabling the stationary probability distribution of the function to be found. But as we are always able to reduce provisioning or dam problems to a general storage problem, this means that a provisioning problem can be interpreted equally well as a dam problem, and vice versa. In other words, any problem and its solution in the one theory will have an exact analogue in the other; equally, any method of attack on a problem in the one theory may be useful in analogous problems of the other.

We proceed to interpret Pitt's results in provisioning theory as solutions in two analogous cases of the infinite dam. We shall also apply Moran's methods in the theory of dams to some problems in the theory of provisioning.

IV. Interpretation of Pitt's results in the case of the infinite dam

Pitt's equation for the stock deficit

$$K - s(t) = c(t) - r(t) \quad (r(t) \leqslant c(t)),$$

was seen to be an example of the storage equation

$$S(t) = I(t) - D(t),$$

where at any time t (t continuous), $S(t)$ is discrete and lies in the range $(0, \infty)$. $I(t)$ is a random discrete input function, and $D(t)$ is a discrete output function depending on $I(t)$, ($D(t) \leqslant I(t)$). Equally, a storage equation of this form can represent the discrete dam content $Z(t)$ ($S(t) = Z(t)$), at any time t, where the dam is infinite ($0 \leqslant Z(t) < \infty$). In this case, interpreting the specific conditions of Pitt's problems of provisioning for an infinite dam, we have that the discrete input function $I(t)$ will be of Poisson type with parameter a such that, in a small interval of time δt, the probabilities that one and no units of water flow into the dam are respectively

$$\begin{aligned}
\Pr\{I(t+\delta t) - I(t) = 1\} &= a\,\delta t, \\
\Pr\{I(t+\delta t) - I(t) = 0\} &= 1 - a\,\delta t,
\end{aligned} \tag{10}$$

and increases in $I(t)$ in non-overlapping intervals are independent. The release rules corresponding to Pitt's two replacement functions are such that the resulting output functions will be:

Type 1.
$$D_1(t) = [I(t-T)\,M^{-1}]\,M; \tag{11}$$

here, releases of M units are made at irregular time intervals t, after a time-lag T from the times $t - T$ when $I(t-T)$ is an integral multiple of M. In practice, this might mean that at every time when the input increased by M units, a decision would be taken to release M units from the dam, but this decision would be carried out after a time-lag T;

Type 2.
$$D_2(t) = I\{[(t-T)\,a M^{-1}]\,M a^{-1}\}; \tag{12}$$

here, releases of an irregular number of units equal to the input in the previous time interval $((k-1)\,M a^{-1}, k M a^{-1})$ are made at regular times $k M a^{-1} + T$ ($k = 1, 2, \ldots$), with time-lag T after the interval. In practice, this might mean that at the regular times $k M a^{-1}$, a decision would be taken to release the input in the previous interval $M a^{-1}$, but this decision would be carried out after a time-lag T.

We proceed to follow Pitt exactly in deriving the stationary distributions of the dam content $Z(t)$ when the release rules of types 1 and 2 are functioning, and in comparing the mean contents of the dam for these cases.

(1) The stationary distribution of Z(t) for release rule of type 1

In this case, where the output function is given by (11), the storage equation for the dam content can be written

$$Z(t) = I(t) - [I(t-T)\,M^{-1}]\,M \quad (0 \leqslant Z(t) < \infty).$$

Let the input function at times t, and $t - T$ be such that

$$I(t-T) = i \quad \text{and} \quad I(t) - I(t-T) = j,$$

where i and j can take the values $0, 1, 2, \ldots$. Then, writing

$$P_n(t) = \Pr\{Z(t) = n\} \quad (n = 0, 1, 2, \ldots),$$

for the probability that the dam content be n units at time t, we have that

$$P_n(t) = \Pr\{i+j-[iM^{-1}]M = n\}.$$

Now suppose that $kM \leqslant i \leqslant (k+1)M-1 \quad (k = 0, 1, \ldots),$
then it follows that

$$[iM^{-1}] = k, \quad i = kM+v \quad \text{and} \quad j = n-v,$$

where v can take the values $0, 1, \ldots, M-1$. From the independence condition for increases in $I(t)$ in non-overlapping intervals, it follows that

$$
\begin{aligned}
P_n(t) &= \sum_{i,j} \Pr\{I(t-T)=i\}.\Pr\{I(t)-I(t-T)=j\} \\
&= \sum_{v=0}^{M-1} \sum_{k=0}^{\infty} e^{-a(t-T)} \frac{\{a(t-T)\}^{kM+v}}{(kM+v)!} g_{n-v},
\end{aligned}
\tag{13}
$$

where we write $\qquad g_r = e^{-aT}\dfrac{(aT)^r}{r!} \quad \text{for} \quad r \geqslant 0,$

and $\qquad\qquad\qquad g_r = 0 \qquad\qquad \text{for} \quad r < 0.$

The stationary distribution is obtained when $t \to \infty$; we shall write it $P_n = \lim\limits_{t\to\infty} P_n(t)$. It follows that

$$
P_n = \sum_{v=0}^{M-1} g_{n-v} \lim_{t\to\infty} \sum_{k=0}^{\infty} e^{-a(t-T)} \frac{\{a(t-T)\}^{kM+v}}{(kM+v)!}.
\tag{14}
$$

Now consider the series $\sum\limits_{k=0}^{\infty} e^{-y}\dfrac{y^{kM+v}}{(kM+v)!}$, where $y = a(t-T)$; if $y = v+(N+1)M$, we have that since $v \leqslant M-1$, then for $k = 0$,

$$
M^{-1}\left\{1+y+\ldots+\frac{y^v}{v!}\right\} < \frac{y^v}{v!} < M^{-1}\left\{\frac{y^v}{v!}+\ldots+\frac{y^{v+M-1}}{(v+M-1)!}\right\}.
\tag{15}
$$

For $k = 1, 2, \ldots, N$, the following inequalities hold:

$$
\begin{aligned}
M^{-1}&\left\{\frac{y^{v+(k-1)M+1}}{(v+(k-1)M+1)!}+\ldots+\frac{y^{v+kM}}{(v+kM)!}\right\} \\
&< \frac{y^{v+kM}}{(v+kM)!} < M^{-1}\left\{\frac{y^{v+kM}}{(v+kM)!}+\ldots+\frac{y^{v+(k+1)M-1}}{(v+(k+1)M-1)!}\right\},
\end{aligned}
\tag{16}
$$

but for $k = N+1$, only one side of the inequality is valid

$$
M^{-1}\left\{\frac{y^{v+NM+1}}{(v+NM+1)!}+\ldots+\frac{y^{v+(N+1)M}}{(v+(N+1)M)!}\right\} < \frac{y^{v+(N+1)M}}{(v+(N+1)M)!}.
\tag{17}
$$

Further, for all values of k greater than $N+1$ $(k = N+2, \ldots)$, the inequalities are reversed

$$
\begin{aligned}
M^{-1}&\left\{\frac{y^{v+kM+1}}{(v+kM+1)!}+\ldots+\frac{y^{v+(k+1)M}}{(v+(k+1)M)!}\right\} \\
&< \frac{y^{v+kM}}{(v+kM)!} < M^{-1}\left\{\frac{y^{v+(k-1)M}}{(v+(k-1)M)!}+\ldots+\frac{y^{v+kM-1}}{(v+kM-1)!}\right\}.
\end{aligned}
\tag{18}
$$

Summing the inequalities (15), (16), (17) and (18), for all values of k, and multiplying the result by e^{-v}, we obtain

$$
M^{-1}\left\{1-\sum_{i=1}^{M} e^{-v}\frac{y^{v+(N+1)M+i}}{(v+(N+1)M+i)!}\right\} < \sum_{k=0}^{\infty} e^{-v}\frac{y^{v+kM}}{(v+kM)!} < M^{-1}\left\{1+e^{-v}\frac{y^{v+(N+1)M}}{(v+(N+1)M)!}\right\}.
$$

Now let $N \to \tau$; y will also tend to infinity, and we have that

$$M^{-1} \leqslant \lim_{y \to \infty} \sum_{k=0}^{\infty} e^{-y} \frac{y^{v+kM}}{(v+kM)!} \leqslant M^{-1},$$

so that the equation (14) for P_n can be written

$$P_n = M^{-1} \sum_{v=0}^{M-1} g_{n-v} \tag{19}$$

The mean content of the dam is given by

$$\mathscr{E}(Z) = M^{-1} \sum_{n=0}^{\infty} n \sum_{v=0}^{M-1} g_{n-v}$$

$$= M^{-1} \left\{ \sum_{n=0}^{\infty} (n+(n+1)+\ldots+(n+M-1)) \right\} g_n$$

$$= aT + \tfrac{1}{2}(M-1). \tag{20}$$

(2) *The stationary distribution of $Z(t)$ for release rule of type 2*

In this case, the output function is (12), and releases are made at regular times $kMa^{-1}+T$ $(k=0,1,\ldots)$. Consider the time interval

$$kMa^{-1}+T \leqslant t < (k+1)Ma^{-1}+T$$

between two releases, where T may have any value $(T \gtrless Ma^{-1})$; at any time in this interval, the content $Z(t)$ of the dam has the value

$$Z(t) = I(t) - I([(t-T)aM^{-1}]Ma^{-1})$$

$$= I(t) - I(kMa^{-1}).$$

Now if we write for the probability that the dam content be n units at time t,

$$P_n(t) = \Pr\{Z(t)=n\} \quad (n=0,1,\ldots),$$

then we have that

$$P_n(t) = \Pr\{I(t) - I(kMa^{-1}) = n\}$$

$$= e^{-a(t-kMa^{-1})} \frac{(at-kM)^n}{n!}. \tag{21}$$

This is periodic in Ma^{-1}, so that in order to obtain the stationary distribution $\{P_n\}$, we write

$$P_n = aM^{-1} \int_{t=kMa^{-1}+T}^{(k+1)Ma^{-1}+T} P_n(t) \, dt$$

$$= M^{-1} \int_{u=0}^{M} e^{-(u+aT)} \frac{(u+aT)^n}{n!} \, du,$$

where

$$u = a(t-kMa^{-1}-T).$$

On expanding the binomial, we have

$$P_n = \sum_{v=0}^{n} e^{-aT} \frac{(aT)^{n-v}}{(n-v)!} M^{-1} \int_0^M e^{-u} \frac{u^v}{v!} \, du$$

$$= \sum_{v=0}^{n} g_{n-v} Q_v, \tag{22}$$

where

$$Q_v = M^{-1} \int_0^M e^{-u} \frac{u^v}{v!} \, du.$$

The mean of the dam content in this case is given by

$$\mathcal{E}(Z) = \sum_{n=0}^{\infty} n \sum_{v=0}^{n} g_{n-v} Q_v$$

$$= \sum_{v=0}^{\infty} Q_v \sum_{n=0}^{\infty} (n+v) g_n$$

$$= \sum_{v=0}^{\infty} Q_v (aT + v).$$

It is easy to see that

$$\sum_{v=0}^{\infty} Q_v = M^{-1} \int_0^M \sum_{v=0}^{\infty} e^{-u} \frac{u^v}{v!} du = 1;$$

$$\sum_{v=0}^{\infty} v Q_v = M^{-1} \int_0^M \sum_{v=0}^{\infty} v e^{-u} \frac{u^v}{v!} du = \tfrac{1}{2} M,$$

whence the mean is obtained as

$$\mathcal{E}(Z) = aT + \tfrac{1}{2} M. \tag{23}$$

Comparing (23) with (20), we note that the second release rule gives a larger mean content of the dam than the first.

V. Application of Moran's methods to provisioning with a finite positive stock

In dealing with provisioning, our notation will be more suggestive of the conditions of the problems if we write the general storage equation in the form

$$S(t) = K - s(t) = c(t) - r(t) - F(t), \tag{24}$$

where $S(t)$ is the stock deficit, $s(t)$ the stock with maximum value K, and $c(t)$, $r(t)$ and $F(t)$, consumption, replacement and overflow functions respectively.

Pitt considered cases where, since components not in stock could be borrowed, the deficit could become infinite, so that for replacement functions always smaller than or equal to consumption functions ($r(t) \leqslant c(t)$), no overflow function was necessary, and $F(t)$ was equal to zero. We shall discuss the case, in practice equally realistic, where the stock is finite and positive ($0 \leqslant s(t) \leqslant K$), so that the deficit is also finite and positive ($0 \leqslant S(t) \leqslant K$). This is the case where no borrowing is allowed, and consumption is lost between times when the stock becomes zero (or the deficit K) until there is a replacement which raises the stock above zero (or decreases the deficit below K). This will require, in cases where $r(t)$ is smaller than $c(t)$, that an overflow function operates so as to restrict the deficit to a value no greater than K, $S(t) \leqslant K$. In these conditions it is difficult to discuss $S(t)$ for all time t (t continuous), though it remains comparatively simple to consider it at specific time intervals. This was the case for the dam content $Z(t)$, defined at times $t = 0, 1, 2, \ldots$, for which Moran framed a set of equations which gave its stationary probability distribution; we proceed to apply Moran's methods to three particular problems of provisioning with finite positive stock.

(1) *Provisioning with replacements at fixed times $t = 1, 2, \ldots$, and no time-lag*

This problem was designed to provide an example of an exact analogue in provisioning theory of the dam problem considered by Moran; in practice, however, it gives rise to a perfectly possible situation. Suppose that at fixed times $t = 1, 2, \ldots$, a consignment of M components arrives at a store for a stock replacement; if at these times there is a deficit equal to or greater than M, the replacement consists of all M components, but if the deficit

is smaller than M, only enough components are added to the stock from the consignment to reduce the deficit to zero. There is no time lag between the evaluation of the deficit at times $t = 0$, and the deliveries of components for replacements at times t. This might well be the case for a truck containing a consignment of M components which, in its regular deliveries at a store, discharges for replacements only as many of the components as may be required. This replacement policy can be expressed by the replacement function $r(t)$, a step function with jumps at times $t = 1, 2, \ldots$, such that

$$r(t) - r(t-1) = M \qquad \text{if} \quad S(t-0) \geqslant M,$$

or

$$r(t) - r(t-1) = S(t-0) \quad \text{if} \quad S(t-0) < M.$$

Let the discrete consumption function $c(t)$ for the store be defined in any interval of time θ by the distribution $\{p_i(\theta)\}$ $(i = 0, 1, 2, \ldots)$, such that

$$\left. \begin{array}{l} (1) \ \ \Pr\{c(t+\theta) - c(t) = i\} = p_i(\theta). \\[4pt] (2) \ \ \text{increases in } c(t) \text{ in non-overlapping time intervals are independent.} \end{array} \right\} \qquad (25)$$

The consumption will be met by components from stock only so long as the stock is positive (or the deficit less than K), but when the stock is zero (or the deficit K), no further orders can be met until a replacement arrives. This condition defines the overflow function $F(t)$, for in the interval of time $(t-1, t)$ between replacements, the overflow required to restrict the deficit to a value K will be such that

$$F(t) - F(t-1) = 0 \quad \text{if} \quad S(t-1) + c(t) - c(t-1) < K,$$

or

$$F(t) - F(t-1) = S(t-1) + c(t) - c(t-1) - K \quad \text{if this is greater than zero.}$$

The storage equation (24) can now be written for times $t = 0, 1, 2, \ldots$, since the consumption, replacement and overflow functions are all defined; this, however, presents no great advantage as it does not lead to a method of solving the problem of finding the stationary probability distribution for the deficit $S(t)$.

If for $t = 0, 1, 2, \ldots$ we write the probabilities that the deficit be i components at times t and $t+1$ respectively as

$$P_i = \Pr\{S(t) = i\}, \quad P_i^{(1)} = \Pr\{S(t+1) = i\},$$

where i takes the values $0, 1, \ldots, K - M$, and if we write $p_i = p_i(1)$ and $q_j = \sum_{i=j}^{\infty} p_i$, then for $K - M \geqslant M$ the equations relating the probabilities are identical to (5), those obtained by Moran in his dam problem. In matrix form, we write $\mathbf{P}^{(1)} = \mathbf{p}\mathbf{P}$, where $\mathbf{P}$, $\mathbf{P}^{(1)}$ are column vectors with elements P_i, $P_i^{(1)}$ respectively, and $\mathbf{p}$ is the matrix of coefficients. Exactly as in Moran's problem, the stationary distribution $P_0, P_1, \ldots, P_{K-M}$ is obtained by solving the equations $\mathbf{P} = \mathbf{p}\mathbf{P}$ together with $\sum_{i=0}^{K-M} P_i = 1$; these $\{P_i\}$ cannot always be obtained in an explicit form, but they can always be evaluated numerically for any given values of the $\{p_i\}$.

(2) Provisioning with replacements at fixed times
$$t = kMa^{-1} + T \ (k = 1, 2, \ldots), \text{ with time-lag } T < Ma^{-1}$$

(a) General equations for the stationary distributions

This problem stems from Pitt's second problem of provisioning under replacement policy of type 2, with replacement function (4). The differences arising are due to the restriction of

the stock $s(t)$ to the range $0 \leqslant s(t) \leqslant K$, or the deficit $S(t)$ to the range $0 \leqslant S(t) \leqslant K$, and to the consideration of the deficit at specific time intervals only. The consumption function $c(t)$ is defined by (25) as in the previous problem, and the replacement function (4) is

$$r_2(t) = c([[(t-T)aM^{-1}]Ma^{-1}),$$

where an irregular number of components equal to the consumption in the previous time interval Ma^{-1} are ordered at regular times kMa^{-1} ($k = 1, 2, \ldots$), and are delivered at times $kMa^{-1}+T$, after a positive time-lag T, where $T < Ma^{-1}$. As the deficit is restricted to values no greater than K, consumption is lost in the intervals after the deficit reaches the value K until a replacement arrives to reduce the deficit to a value less than K. This defines the overflow function in such a way that, in the intervals kMa^{-1}, $kMa^{-1}+T$, between orders and deliveries, the overflow is

$$F(kMa^{-1}+T) - F(kMa^{-1}) = 0 \quad \text{if} \quad S(kMa^{-1}) + c(kMa^{-1}+T) - c(kMa^{-1}) \leqslant K$$

or

$$F(kMa^{-1}+T) - F(kMa^{-1})$$
$$= S(kMa^{-1}) + c(kMa^{-1}+T) - c(kMa^{-1}) - K \quad \text{if this is greater than zero.} \tag{26}$$

Similarly, in the intervals $kMa^{-1}+T$, $(k+1)Ma^{-1}$, between deliveries and new orders, the overflow is

$$F((k+1)Ma^{-1}) - F(kMa^{-1}+T) = 0 \quad \text{if} \quad S(kMa^{-1}+T) + c((k+1)Ma^{-1}) - c(kMa^{-1}+T) \leqslant K$$

or

$$F((k+1)Ma^{-1}) - F(kMa^{-1}+T)$$
$$= S(kMa^{-1}+T) + c((k+1)Ma^{-1}) - c(kMa^{-1}+T) - K \quad \text{if this is greater than zero.}$$

The storage equation (24) is now fully defined at times kMa^{-1}, $kMa^{-1}+T$, but, as in the previous problem, this does not lead to a solution of the stationary distribution of $S(t)$, the deficit. We can, however, obtain sets of equations relating the probabilities

$$P_0, P_1, P_2, \ldots, P_K, \quad P_0^{(1)}, P_1^{(1)}, P_2^{(1)}, \ldots, P_K^{(1)}, \quad \text{and} \quad P_0^{(2)}, P_1^{(2)}, P_2^{(2)}, \ldots, P_K^{(2)},$$

that the deficit $S(t)$ be equal to $0, 1, 2, \ldots, K$, at the times kMa^{-1}, $kMa^{-1}+T$, $(k+1)Ma^{-1}$ respectively. Writing $p_i = p_i(T)$, and $q_j = \sum_{i=j}^{\infty} p_i$, we have for the times kMa^{-1}, $kMa^{-1}+T$, the relations

$$P_0^{(1)} = p_0(P_0+P_1+\ldots+P_{K-1}) + q_0 P_K,$$
$$P_1^{(1)} = p_1(P_0+P_1+\ldots+P_{K-2}) + q_1 P_{K-1},$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$P_K^{(1)} = q_K P_0, \tag{27}$$

or in matrix form, $\mathbf{P}^{(1)} = \mathbf{p}\mathbf{P}$:

$$\begin{pmatrix} P_0^{(1)} \\ P_1^{(1)} \\ P_2^{(1)} \\ \vdots \\ P_K^{(1)} \end{pmatrix} = \begin{pmatrix} p_0 & p_0 & \cdots & \cdots & \cdots & p_0 & q_0 \\ p_1 & p_1 & \cdots & \cdots & p_1 & q_1 & 0 \\ p_2 & p_2 & \cdots & p_2 & q_2 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ q_K & 0 & \cdots & \cdots & \cdots & \cdots & 0 \end{pmatrix} \begin{pmatrix} P_0 \\ P_1 \\ P_2 \\ \vdots \\ P_K \end{pmatrix}. \tag{28}$$

Similarly, writing $p_i^1 = p_i(Ma^{-1} - T)$ and $q^1 = \sum_{i=i}^{\infty} p_i^{(1)}$, we have for the times $kMa^{-1} \cdot T$, $(k+1)Ma^{-1}$, the relations

$$
\begin{aligned}
P_0^{(2)} &= p_0^{(1)} P_0^{(1)}, \\
P_1^{(2)} &= p_1^{(1)} P_0^{(1)} + p_0^{(1)} P_1^{(1)}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
P_K^{(2)} &= q_K^{(1)} P_0^{(1)} + q_{K-1}^{(1)} P_1^{(1)} + \ldots + q_0^{(1)} P_K^{(1)},
\end{aligned}
\tag{29}
$$

or in matrix form, $\mathbf{P}^{(2)} = \mathbf{p}^{(1)} \mathbf{P}^{(1)}$:

$$
\begin{pmatrix} P_0^{(2)} \\ P_1^{(2)} \\ P_2^{(2)} \\ \vdots \\ P_K^{(2)} \end{pmatrix} = \begin{pmatrix} p_0^{(1)} & 0 & 0 & \cdots & \cdots & 0 \\ p_1^{(1)} & p_0^{(1)} & 0 & \cdots & \cdots & 0 \\ p_2^{(1)} & p_1^{(1)} & p_0^{(1)} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ q_K^{(1)} & q_{K-1}^{(1)} & q_{K-2}^{(1)} & \cdots & \cdots & q_0^{(1)} \end{pmatrix} \begin{pmatrix} P_0^{(1)} \\ P_1^{(1)} \\ P_2^{(1)} \\ \vdots \\ P_K^{(1)} \end{pmatrix}.
\tag{30}
$$

The relations between $\mathbf{P}^{(2)}$ and $\mathbf{P}$, the probabilities of $S(t)$ at consecutive ordering times kMa^{-1}, $(k+1)Ma^{-1}$, are then given by

$$
\mathbf{P}^{(2)} = \mathbf{p}^{(1)}\mathbf{p}\mathbf{P},
\tag{31}
$$

where the matrix $\mathbf{p}^{(1)}\mathbf{p}$ is square. For ordering times, the stationary probability distribution $P_0, P_1, \ldots, P_K$ can be found by putting $P_i^{(2)} = P_i$ for $i = 0, 1, \ldots, K$, and solving the set of equations $\mathbf{P} = \mathbf{p}^{(1)}\mathbf{p}\mathbf{P}$ together with $\sum_{i=0}^{K} P_i = 1$. This solution may not be obtained easily in an explicit form, but it can always be computed numerically for any given distribution $\{p_i(\theta)\}$, $i = 0, 1, \ldots$. We note, in passing, that this stationary distribution for ordering times will be easier to compute if we first obtain the stationary probability distribution $P_0^{(1)}, P_1^{(1)}, \ldots, P_K^{(1)}$, for delivery times. Writing $P_i^{(3)} = \Pr\{S((k+1)Ma^{-1}+T) = i\}$ for the probability that $S(t)$ have a value i at delivery time $(k+1)Ma^{-1}+T$, we have that the relation between $\mathbf{P}^{(3)}$ and $\mathbf{P}^{(1)}$, the probabilities of $S(t)$ at consecutive delivery times $kMa^{-1}+T$, $(k+1)Ma^{-1}+T$, is given by

$$
\mathbf{P}^{(3)} = \mathbf{p}\mathbf{p}^{(1)}\mathbf{P}^{(1)},
\tag{32}
$$

where the matrix $\mathbf{p}\mathbf{p}^{(1)}$ is triangular. The stationary probability distribution for delivery times, $P_0^{(1)}, P_1^{(1)}, \ldots, P_K^{(1)}$, can be found by putting $P_i^{(3)} = P_i^{(1)}$ for $i = 0, 1, \ldots, K$, and solving the equations $\mathbf{P}^{(1)} = \mathbf{p}\mathbf{p}^{(1)}\mathbf{P}^{(1)}$, together with $\sum_{i=0}^{K} P_i^{(1)} = 1$. For numerical computation of $\mathbf{P}^{(1)}$, the work is considerably lightened by the fact that $\mathbf{p}\mathbf{p}^{(1)}$ is triangular; once the stationary probability distribution for delivery times is obtained, the simple relation $\mathbf{P} = \mathbf{p}^{(1)}\mathbf{P}^{(1)}$ will enable the stationary distribution for ordering times, $P_0, P_1, \ldots, P_K$ to be found.

(b) A numerical example

As an illustration of the computations necessary to evaluate the stationary distributions $\{P_i\}$ and $\{P_i^{(1)}\}$, for ordering and delivery times, we construct an example in which the consumption function $c(t)$ has, associated with it, a Poisson distribution such that

$$
p_i(\theta) = e^{-a\theta} \frac{(a\theta)^i}{i!} \quad (i = 0, 1, \ldots).
$$

We choose the maximum deficit $K = 9$, the time-lag between order and delivery $T = \frac{1}{2}$, the time interval between orders $Ma^{-1} = 1$, and the mean consumption per unit time $a = 4$. Then, since $T = Ma^{-1} - T = \frac{1}{2}$, we have that the p_i and $p_i^{(1)}$ defined previously are equal,

$$p_i = p_i^{(1)} = e^{-2}2^i/i! \quad (i = 0, 1, \ldots).$$

From Molina's tables for *Poisson's Exponential Binomial Limit*, we obtain

$$
\begin{aligned}
&p_0 = p_0^{(1} = 0{\cdot}135335, \quad q_0 = q_0^{(1)} = 1{\cdot}000000, \\
&p_1 = p_1^{(1)} = 0{\cdot}270671, \quad q_1 = q_1^{(1)} = 0{\cdot}864665, \\
&p_2 = p_2^{(1)} = 0{\cdot}270671, \quad q_2 = q_2^{(1)} = 0{\cdot}593994, \\
&p_3 = p_3^{(1)} = 0{\cdot}180447, \quad q_3 = q_3^{(1)} = 0{\cdot}323324, \\
&p_4 = p_4^{(1)} = 0{\cdot}090224, \quad q_4 = q_4^{(1)} = 0{\cdot}142877, \\
&p_5 = p_5^{(1)} = 0{\cdot}036089, \quad q_5 = q_5^{(1)} = 0{\cdot}052653, \\
&p_6 = p_6^{(1)} = 0{\cdot}012030, \quad q_6 = q_6^{(1)} = 0{\cdot}016564, \\
&p_7 = p_7^{(1)} = 0{\cdot}003437, \quad q_7 = q_7^{(1)} = 0{\cdot}004534, \\
&p_8 = p_8^{(1)} = 0{\cdot}000859, \quad q_8 = q_8^{(1)} = 0{\cdot}001097, \\
&\phantom{p_8 = p_8^{(1)} = 0{\cdot}000859,} \quad q_9 = q_9^{(1)} = 0{\cdot}000237.
\end{aligned}
\tag{33}
$$

The equations giving the stationary distribution $\{P_i^{(1)}\}$ for delivery times are $\mathbf{P}^{(1)} = \mathbf{pp}^{(1)}\mathbf{P}$, together with $\sum_{i=0}^{K} P_i^{(1)} = 1$; these, in the case of the $\{p_i\}$ and $\{q_i\}$ given in (33), result in eleven equations, not all independent, for ten unknowns:

$$-0{\cdot}864460P_0^{(1)} + 0{\cdot}136284P_1^{(1)} + 0{\cdot}139256P_2^{(1)} + 0{\cdot}149657P_3^{(1)} + 0{\cdot}180862P_4^{(1)}$$
$$+ 0{\cdot}258876P_5^{(1)} + 0{\cdot}414902P_6^{(1)} + 0{\cdot}645941P_7^{(1)} + 0{\cdot}882981P_8^{(1)} + P_9^{(1)} = 0,$$

$$0{\cdot}271117P_0^{(1)} - 0{\cdot}727584P_1^{(1)} + 0{\cdot}276590P_2^{(1)} + 0{\cdot}287625P_3^{(1)} + 0{\cdot}310012P_4^{(1)}$$
$$+ 0{\cdot}339183P_5^{(1)} + 0{\cdot}343934P_6^{(1)} + 0{\cdot}270671P_7^{(1)} + 0{\cdot}117019P_8^{(1)} \qquad = 0,$$

$$0{\cdot}271486P_0^{(1)} + 0{\cdot}273334P_1^{(1)} - 0{\cdot}722144P_2^{(1)} + 0{\cdot}285591P_3^{(1)} + 0{\cdot}290341P_4^{(1)}$$
$$+ 0{\cdot}270671P_5^{(1)} + 0{\cdot}197408P_6^{(1)} + 0{\cdot}080388P_7^{(1)} \qquad = 0,$$

$$0{\cdot}181348P_0^{(1)} + 0{\cdot}182615P_1^{(1)} + 0{\cdot}183837P_2^{(1)} - 0{\cdot}819553P_3^{(1)} + 0{\cdot}160777P_4^{(1)}$$
$$+ 0{\cdot}111934P_5^{(1)} + 0{\cdot}043756P_6^{(1)} \qquad = 0,$$

$$0{\cdot}090630P_0^{(1)} + 0{\cdot}090224P_1^{(1)} + 0{\cdot}086834P_2^{(1)} + 0{\cdot}075304P_3^{(1)} - 0{\cdot}949117P_4^{(1)}$$
$$+ 0{\cdot}019336P_5^{(1)} \qquad = 0,$$

$$0{\cdot}035683P_0^{(1)} + 0{\cdot}033922P_1^{(1)} + 0{\cdot}028904P_2^{(1)} + 0{\cdot}019136P_3^{(1)} + 0{\cdot}007126P_4^{(1)}$$
$$- P_5^{(1)} \qquad = 0,$$

$$0{\cdot}011129P_0^{(1)} + 0{\cdot}009368P_1^{(1)} + 0{\cdot}006111P_2^{(1)} + 0{\cdot}002242P_3^{(1)}$$
$$- P_6^{(1)} \qquad = 0,$$

$$0{\cdot}002623P_0^{(1)} + 0{\cdot}001692P_1^{(1)} + 0{\cdot}000614P_2^{(1)} \qquad = 0,$$

$$0{\cdot}000413P_0^{(1)} + 0{\cdot}000148P_1^{(1)} \qquad - P_7^{(1)} \qquad = 0,$$

$$0{\cdot}000032P_0^{(1)} \qquad - P_8^{(1)} \qquad = 0,$$

$$P_0^{(1)} + \quad P_1^{(1)} + \quad P_2^{(1)} + \quad P_3^{(1)} + \quad P_4^{(1)} \qquad - P_9^{(1)} = 0,$$
$$+ \quad P_5^{(1)} + \quad P_6^{(1)} + \quad P_7^{(1)} + \quad P_8^{(1)} + P_9^{(1)} = 1.$$

Solving these equations by straightforward elimination which is greatly simplified by the fact that pp^1 is triangular, we obtain the following values for $P_i^{(1)}$:

$$P_0^{(1)} = 0.148685, \quad P_5^{(1)} = 0.026833,$$
$$P_1^{(1)} = 0.281325, \quad P_6^{(1)} = 0.006382,$$
$$P_2^{(1)} = 0.277080, \quad P_7^{(1)} = 0.001036, \tag{34}$$
$$P_3^{(1)} = 0.177621, \quad P_8^{(1)} = 0.000103,$$
$$P_4^{(1)} = 0.080930, \quad P_9^{(1)} = 0.000005.$$

From these we can compute the stationary distribution $\{P_i\}$ for ordering times, using the relations between P and $P^{(1)}$ which are given in matrix form by $P = p^{(1)}P^{(1)}$. This is easier than solving directly the equations $P = p^{(1)}pP$ and $\sum_{i=0}^{K} P_i = 1$, where the matrix $p^{(1)}p$ is square. We obtain for $\{P_i\}$ the values:

$$P_0 = 0.020122, \quad P_5 = 0.154360,$$
$$P_1 = 0.078318, \quad P_6 = 0.099024,$$
$$P_2 = 0.153890, \quad P_7 = 0.053655,$$
$$P_3 = 0.202012, \quad P_8 = 0.025004,$$
$$P_4 = 0.198206, \quad P_9 = 0.015410.$$

(c) *Comparison with Pitt's results for an infinite deficit*

It is interesting to compare the stationary distribution of the deficit $S(t)$ in the present case, where it is finite and restricted to the range $0 \leqslant S(t) \leqslant K$, with that obtained by Pitt for an infinite deficit with the same types of consumption and replacement functions. Pitt's problem for the infinite deficit, when the consumption function is of Poisson type, and the replacement function is the $r_2(t)$ defined in (4), is the exact analogue of the problem of the infinite dam considered in §IV (2). With a slight change in the notation of (22) to avoid confusion, we write the stationary distribution of the deficit $S(t)$ when its range is $0 \leqslant S(t) < \infty$ as $\{\Pi_i\}$ ($i = 0, 1, \ldots$), where

$$\Pi_i = \sum_{v=0}^{i} p_{i-v} Q_v. \tag{35}$$

Here, the $p_{i-v} = e^{-aT} \dfrac{(aT)^{i-v}}{(i-v)!}$, and the Q_v are given by

$$Q_v = M^{-1} \int_0^M e^{-u} \frac{u^v}{v!} \, du$$

$$= M^{-1} \sum_{i=v+1}^{\infty} e^{-M} \frac{M^i}{i!}.$$

We proceed to obtain a similar stationary distribution, also averaged over the time interval Ma^{-1}, for the present case of the finite deficit, and we shall write it $\{\Pi_i^{(1)}\}$ ($i = 0, 1, \ldots, K$). To do this we consider the time interval θ, where $\theta \leqslant Ma^{-1}$ such that $kMa^{-1} + T + \theta$ is any time between two deliveries at times $kMa^{-1} + T$, and $(k+1)Ma^{-1} + T$. Let the probabilities that there be consumption i in this interval be

$$p_i(\theta) = e^{-a\theta} \frac{(a\theta)^i}{i!},$$

then if $\{P_i(kMa^{-1}+T+\theta)\}$ represent the distribution of the deficit $S(t)$ at the time $kMa^{-1}+T+\theta$, this will be given in matrix form by

$$
\begin{pmatrix}
P_0(kMa^{-1}+T+\theta) \\
P_1(kMa^{-1}+T+\theta) \\
\cdots\cdots\cdots\cdots \\
P_{K-1}(kMa^{-1}+T+\theta) \\
P_K(kMa^{-1}+T+\theta)
\end{pmatrix}
=
\begin{pmatrix}
p_0(\theta) & 0 & \cdots & \cdots & \cdots & 0 \\
p_1(\theta) & p_0(\theta) & 0 & \cdots & & 0 \\
\cdots & & \cdots & \cdots & \cdots & \cdots \\
p_{K-1}(\theta) & p_{K-2}(\theta) & \cdots & \cdots & p_0(\theta) & 0 \\
q_K(\theta) & q_{K-1}(\theta) & \cdots & \cdots & q_1(\theta) & q_0(\theta)
\end{pmatrix}
\begin{pmatrix}
P_0^{(1)} \\
P_1^{(1)} \\
\vdots \\
P_{K-1}^{(1)} \\
P_K^{(1)}
\end{pmatrix},
$$

where $q_j(\theta) = \sum\limits_{i=j}^{\infty} p_i(\theta)$, and $P_0^{(1)}, P_1^{(1)}, ..., P_K^{(1)}$, are the stationary probability distribution of the deficit at delivery times. We can write these in the forms

$$
P_i(kMa^{-1}+T+\theta) = \sum_{v=0}^{i} P_{i-v}^{(1)} p_v(\theta) \quad (i = 0, 1, ..., K-1),
$$

$$
P_K(kMa^{-1}+T+\theta) = \sum_{v=0}^{K} P_{K-v}^{(1)} q_v(\theta).
$$

The stationary distribution $\{\Pi_i^{(1)}\}$ will then be obtained from the $\{P_i(kMa^{-1}+T+\theta)\}$ by averaging them over the interval of time Ma^{-1}, so that

$$
\left.
\begin{aligned}
\Pi_i^{(1)} &= aM^{-1} \int_0^{Ma^{-1}} \sum_{v=0}^{i} P_{i-v}^{(1)} p_v(\theta)\, d\theta, \\
&= \sum_{v=0}^{i} P_{i-v}^{(1)} Q_v \quad (i = 0, 1, ..., K-1); \\
\Pi_K^{(1)} &= aM^{-1} \int_0^{Ma^{-1}} \sum_{v=0}^{K} P_{K-v}^{(1)} q_v(\theta)\, d\theta, \\
&= \sum_{v=0}^{K} P_{K-v}^{(1)} \sum_{j=v}^{\infty} Q_j,
\end{aligned}
\right\}
\tag{36}
$$

where there is some similarity of form with (35).

In the case of our numerical example (§ V (2)(b)), where $K = 9$, $T = \frac{1}{2}$, $M = a = 4$, the values of the p_{i-v} in (35) are given by equation (33) for values of $i = 0, 1, ..., 8$; the values of the Q_v in (35) and (36) are given for $v = 0, 1, ..., 8$ by

$$
\begin{aligned}
&Q_0 = 0{\cdot}245421, &&Q_3 = 0{\cdot}141633, &&Q_6 = 0{\cdot}027669, \\
&Q_1 = 0{\cdot}227106, &&Q_4 = 0{\cdot}092791, &&Q_7 = 0{\cdot}012784, \\
&Q_2 = 0{\cdot}190474, &&Q_5 = 0{\cdot}053718, &&Q_8 = 0{\cdot}005341;
\end{aligned}
$$

and the values of the $\{P_i^{(1)}\}$ are given by (34). Using these, we compute

$$
\begin{aligned}
&\Pi_0 = 0{\cdot}033214, &&\Pi_0^{(1)} = 0{\cdot}036490, \\
&\Pi_1 = 0{\cdot}097164, &&\Pi_1^{(1)} = 0{\cdot}102810, \\
&\Pi_2 = 0{\cdot}153677, &&\Pi_2^{(1)} = 0{\cdot}160212, \\
&\Pi_3 = 0{\cdot}176480, &&\Pi_3^{(1)} = 0{\cdot}181162, \\
&\Pi_4 = 0{\cdot}165573, &&\Pi_4^{(1)} = 0{\cdot}166619, \\
&\Pi_5 = 0{\cdot}134440, &&\Pi_5^{(1)} = 0{\cdot}132132, \\
&\Pi_6 = 0{\cdot}097291, &&\Pi_6^{(1)} = 0{\cdot}093169, \\
&\Pi_7 = 0{\cdot}063731, &&\Pi_7^{(1)} = 0{\cdot}059328, \\
&\Pi_8 = 0{\cdot}038132, &&\Pi_8^{(1)} = 0{\cdot}034385, \\
&\sum_{i=9}^{\infty} \Pi_i = 0{\cdot}040298, &&\Pi_9^{(1)} = 0{\cdot}033693.
\end{aligned}
$$

There would appear to be a close similarity in the values of the $\{\Pi_i\}$ and $\{\Pi_i^{(1)}\}$, so that we could use the more easily computed values of Π_i as approximations for the $\Pi_i^{(1)}$; however, it is not possible without further work to state within what bounds this approximation would be valid.

(3) *Provisioning with orders at fixed times $t = kMa^{-1}$, and replacements at fixed times $t = (k+1)Ma^{-1}$ ($k = 1, 2, ...$), with time-lag Ma^{-1}*

(a) *General equations, and an exact solution for the case of consumption with a discrete geometric distribution*

This is a special case of the previous problem (§ V (2)(a)), when the times for delivery of the replacements and for ordering coincide: that is, a new order for replacements is put in every time a delivery occurs, at time intervals Ma^{-1}. The consumption function is of the same form (25), the replacement function is $r_2(t)$, (4), the overflow function is defined by (26), where the time-lag T is Ma^{-1}: in these conditions, with the same notation as before, the equations (28) relate the probability distributions $\{P_i\}$ and $\{P_i^{(1)}\}$ for $S(t)$ at consecutive ordering and replacement times kMa^{-1}, and $(k+1)Ma^{-1}$ ($k = 1, 2, ...$). To obtain the stationary probability distribution $P_0, P_1, ..., P_K$, for these times, we write $P_i^{(1)} = P_i$ ($i = 0, 1, ..., K$), and solve the equations $\mathbf{P} = p\mathbf{P}$ together with $\sum_{i=0}^{K} P_i = 1$.

If, for convenience, we group these equations in a slightly different way in the top half, we have

$$\left.\begin{aligned}
P_0 &= p_0(P_0 + ... + P_K) + q_1 P_K, \\
P_1 &= p_1(P_0 + ... + P_{K-1}) + q_2 P_{K-1}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
P_{K-1} &= p_{K-1} P_0 + q_{K-1} P_1, \\
P_K &= q_K P_0, \\
1 &= P_0 + P_1 + ... + P_K,
\end{aligned}\right\} \tag{37}$$

which can be solved in pairs, starting with P_0 and P_K, and working towards the centre of the group in the order P_1, P_{K-1}; P_2, P_{K-2}; etc. We obtain the recurrence relations

$$\begin{aligned}
P_0 &= p_0(1 - q_1 q_K)^{-1}, \\
P_K &= q_K P_0 = q_K p_0(1 - q_1 q_K)^{-1}; \\
P_1 &= \{p_1(1 - P_K) + q_2 p_{K-1} P_0\}(1 - q_2 q_{K-1})^{-1}, \\
P_{K-1} &= p_{K-1} P_0 + q_{K-1} P_1;
\end{aligned}$$

for the first two pairs of solutions. Assuming that all probabilities P_n from $P_0, P_1, ...$ to P_{i-1}, and $P_K, P_{K-1}, ...$ to P_{K-i+1} are known, the following formulae enable all other P_i to be found from these by a repetition of the process:

$$\left.\begin{aligned}
P_i &= \left\{p_i\left(1 - \sum_{n=K-i+1}^{K} P_n\right) + q_{i+1} p_{K-i} \sum_{n=0}^{i-1} P_n\right\}(1 - q_{i+1} q_{K-i})^{-1}, \\
P_{K-i} &= p_{K-i} \sum_{n=0}^{i-1} P_n + q_{K-i} P_i \quad (i = 1, ..., K-1).
\end{aligned}\right\} \tag{38}$$

A distribution $\{p_i\}$ which gives a simple explicit solution of these equations is the geometric; it is important to note, however, that this does not satisfy the independence condition for non-overlapping intervals of time postulated in (25). For the geometric distribution

$$p_i = AB^i \quad (i = 0, 1, ...), \tag{39}$$

where $A + B = 1$, we find from the recurrence relations (38) that the stationary distribution of the deficit $\{P_i\}$ is given by

$$P_i = AB^i(1 - B^{K+1})^{-1} \quad (i = 0, 1, \ldots, K).$$

(40)

(b) An exact solution for the case of consumption with a continuous exponential distribution

Suppose that instead of dealing with integral units of the deficit, consumption and replacement, we deal with units of magnitude Δ; Δ will tend to zero, and K and i will tend to infinity in such a way that the maximum deficit $K\Delta$ tends to the value k, and a deficit, consumption or replacement $i\Delta$ tends to the value s $(i = 0, 1, \ldots, K)$. The distribution of the consumption then becomes continuous; for if we put

$$A = 1 - e^{-\mu\Delta},$$
$$B = e^{-\mu\Delta},$$

and let $\Delta \to 0$ in (39), so that $i\Delta \to s$, we have that the distribution of the consumption in the interval of time Ma^{-1} is

$$
\begin{aligned}
p(s)\,ds &= \lim_{\Delta \to 0} \Delta^{-1} p_i\,ds \\
&= \lim_{\Delta \to 0} \Delta^{-1} e^{-\mu i\Delta}(1 - e^{-\mu\Delta})\,ds \\
&= \mu\, e^{-\mu s}\,ds,
\end{aligned}
$$

which is continuous, and of the exponential form. The solutions (40) of equations (37) will lead by the same limiting process to an exact solution of the following provisioning problem.

(41)

A deficit $S(t)$ is defined continuously in the range $0 \leqslant S(t) \leqslant k$, at the fixed times $n.Ma^{-1}$ $(n = 0, 1, \ldots)$, and is subject to the following conditions:

(1) a consumption function $c(t)$ such that in a time interval Ma^{-1} the distribution of the consumption is continuous and of the exponential type

$$
\begin{aligned}
p(s)\,ds &= \Pr\{s < c((n+1)\,Ma^{-1}) - c(nMa^{-1}) < s + ds\} \\
&= \mu\,e^{-\mu s}\,ds;
\end{aligned}
$$

(2) a replacement function such that

$$r_2((n+1)\,Ma^{-1}) = c(nMa^{-1}),$$

where irregular quantities lying between the values $(s, s + ds)$ equal to the consumption in the previous time interval Ma^{-1} are ordered at times $n\,Ma^{-1}$ and delivered and added to the stock at times $(n+1)\,Ma^{-1}$;

(3) an overflow function such that in any time interval Ma^{-1} between ordering and delivery times $n.Ma^{-1}$, $(n+1)\,Ma^{-1}$, the consumption is lost at any time after the deficit rises to the value k, and is otherwise met from the stock.

For small Δ, the equations (37) hold, with the proviso that $\{p_i\}$ represent the probabilities that the consumption in an interval Ma^{-1} is $i\Delta$, and $\{P_i\}$ the stationary probabilities that the deficit have a value $i\Delta$, so that the solutions (40) to them also hold. In the limit as Δ tends to zero in the manner defined above, the continuous stationary distribution of the deficit $S(t)$ is given by

$$
\begin{aligned}
f(s)\,ds &= \lim_{\Delta \to \infty} \Delta^{-1} P_i\,ds \\
&= \mu\,e^{-\mu s}(1 - e^{-\mu k})^{-1}\,ds \quad (0 \leqslant s \leqslant k),
\end{aligned}
$$

a truncated exponential distribution.

(42)

This solution can also be obtained by taking the limit as Δ tends to zero for equations (37); this gives an integral equation for the continuous stationary probability distribution of the deficit,

$$f(s)\,ds = p(s)\,ds \int_0^{k-s} f(u)\,du + f(k-s)\,ds \int_s^k p(u)\,du. \tag{43}$$

It is easily verified that (42) will satisfy this equation.

VI. An exact solution of the storage problem with a Poisson input

In attempting to find algebraic solutions for the stationary probability distribution $\{P_i\}$ of the dam content $Z(t)$ in Moran's problem (§ II), where $\mathbf{P} = p\mathbf{P}$, $\mathbf{P}$ being the column vector of the P_i, it is found that no simple solution in an explicit form exists for an input function of the Poisson type with parameter a, such that in the interval of time $t, t+1$, the probabilities of an input $X(t)$ being i units is

$$p_i = e^{-a} a^i / i! \quad (i = 0, 1, \dots).$$

However, if in this problem the conditions for the input of Poisson type and for the overflow are left unchanged, but the release rule is altered so that instead of the release rule (3) of § II (M units at times $t+1$ if the total content of the dam $Z(t) + X(t)$ is greater or equal to M, or the total content of the dam $Z(t) + X(t)$ if this is less than M), a new release rule is prescribed so that there is an output with a steady flow when the dam contains water, and no output when it is empty, then $Z(t)$ is continuous, and it is possible to obtain solutions in an explicit form for its stationary distribution.

Our method of approach will be to consider Moran's set of equations (5) for an infinitesimal interval of time δt, instead of a unit interval of time; the input will remain of Poisson type, and be a multiple of a definite discrete unit which will not tend to zero, while a small discrete release will be made at the end of each infinitesimal interval of time under a rule of the same type as rule (3) of § II. The equations can then be solved explicitly, and when $\delta t \to 0$, these solutions will provide the continuous stationary probability distribution for the dam content $Z(t)$, which is then itself continuous, when the release rule prescribed gives an output with a steady flow when the dam contains water, or no output when the dam is empty.

Before proceeding further, it is preferable, since the theory of dams was identified with the general storage theory, to view the problem as one of the general storage function. It then becomes possible to interpret the problem as one of provisioning as well; here we would have a consumption of Poisson type, a replacement rule such that there is continuous delivery at a steady rate when the stock is not at its maximum value, and zero when it is, and an overflow function such that the consumption is lost when there is no stock to meet it. In practice, this could conceivably be the case of a stock of grain in a silo, where the replacement is a steady flow of grain which is stopped when the silo is full, and the consumption of Poisson type is lost when the silo is empty, and met from the stock otherwise.

We frame the conditions of the problem for the storage function as follows: a discrete storage function $S(t)$ is defined in the range $0 \leqslant S(t) \leqslant (K-1)\Delta$, where $K = bH + U$ (b, H, U integers) at the fixed times $t = 0, \delta t, 2\delta t, \dots$, and is subject to the conditions that

(1) it is fed by a discrete input function $I(t)$ of Poisson type with parameter a, such that in a small interval of time δt there may be inputs of $H\Delta$ or of no units respectively with probabilities

$$p = \Pr\{I(t+\delta t) - I(t) = H\Delta\} = 1 - e^{-a\delta t},$$
$$q = \Pr\{I(t+\delta t) - I(t) = 0\} = e^{-a\delta t};$$

(2) the output function is such that at fixed times $(n+1)\delta t$ $(n = 0, 1, \ldots)$, the discrete output consists of a quantity $\Delta = C\delta t$ units, where C is a constant, provided that the storage function at a time just previous to $(n+1)\delta t$ is greater than Δ,

$$S(n\delta t) + I((n+1)\delta t) - I(n\delta t) \geqslant \Delta,$$

but is zero otherwise;

(3) the overflow function is such that in the interval $n\delta t, (n+1)\delta t$ there is no overflow if

$$S(n\delta t) + I((n+1)\delta t) - I(n\delta t) \leqslant K\Delta,$$

but there is an overflow

$$S(n\delta t) + I((n+1)\delta t) - I(n\delta t) - K\Delta,$$

if this is greater than zero.

Writing P_i, $P_i^{(1)}$ for the probabilities that $S(t)$ take the values $i\Delta$ at times $n\delta t, (n-1)\delta t$ respectively, where $i = 0, 1, \ldots, K-1$, we have that the relations between these probabilities are

$$
\begin{aligned}
P_0^{(1)} &= qP_0 + qP_1, \\
P_1^{(1)} &= qP_2, \\
&\cdots\cdots\cdots\cdots\cdots \\
P_{H-2}^{(1)} &= qP_{H-1}, \\
P_{H-1}^{(1)} &= pP_0 \\
P_H^{(1)} &= pP_1 +qP_H, \\
&\cdots\cdots\cdots\cdots\cdots \phantom{+qP_{H+1}} +qP_{H+1}, \\
P_{2H-2}^{(1)} &= pP_{H-1} +qP_{2H-1}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
P_{jH-1}^{(1)} &= pP_{(j-1)H} +qP_{jH}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
P_{(j+1)H-2}^{(1)} &= pP_{jH-1} +qP_{(j-1)H-1}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
P_{bH-1}^{(1)} &= pP_{(b-1)H} +qP_{bH}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
P_{K-1}^{(1)} &= P_{bH+U-1}^{(1)} = p(P_{(b-1)H+U} + \ldots + P_{bH+U-1}),
\end{aligned}
\tag{44}
$$

or in matrix form, $\mathbf{P}^{(1)} = \mathbf{p}\mathbf{P}$, where $\mathbf{P}$, $\mathbf{P}^{(1)}$ are column vectors with elements P_i, $P_i^{(1)}$ respectively, and $\mathbf{p}$ is the matrix of coefficients.

Formally, this is Moran's set of equations (5), in which we have set $p_0 = q$, $p_H = p$, and all other $p_i = 0$ for i not equal to 0 or H. For the stationary distribution of $S(t)$, we put $P_i^{(1)} = P_i$, and solve the equations $\mathbf{P} = \mathbf{p}\mathbf{P}$, together with $\sum_{i=0}^{K-1} P_i = 1$. We see clearly that the solution of these equations will fall into the $(b+1)$ distinct classes:

(0)th class: $P_1, P_2, \ldots, P_{H-1}$;

(1)th class: $P_H, P_{H+1}, \ldots, P_{2H-1}$;

$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$

(j)th class: $P_{jH}, P_{jH+1}, \ldots, P_{(j+1)H-1}$;

$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$

(b)th class: $P_{bH}, P_{bH+1}, \ldots, P_{K-1}$.

The solution for the 0th class is immediately obtained from (44)

$$P_R = pP_0 q^{-R} \quad (R = 1, 2, ..., H-1);$$ (45)

for the jth class we obtain from our equations (44) the equation

$$P_{jH-R} + q^{...}\left[P_{...} - p\sum_{...} r^{...}F_{...} + ...\right] \quad (j = 1, ..., \; R = 0, ..., H-1)$$ (46)

creating any probability $P_{...}$ in the jth class with the probabilities $P_{...}, ..., P_{...}$ of the $(j-1)$th class, so that starting with the solution (45) for the 0th class we can obtain solutions progressively for one class after another.

Consider the formula

$$P_{jH-R} = pP_0 q^{...}\left(\sum_{...}(-1)^r p^{...}q^{...}\left[\binom{(j-v)H+R...-1}{v-1} + p\binom{(j-v)H+R+...-1}{v}\right]\right)$$ (47)

where $R = 1, 2, ..., H-1$ for $j = 0$ and $R = 0, 1, ..., H-1$ for $j = 1, 2, ...$ and where it is understood that $\binom{n}{r} = 0$ if $r < 0$ or if $n < r$. For $j = 0$ this gives

$$P_R = pP_0 q^{-R} \quad (R = 1, 2, ..., H-1),$$

which is the solution (45) for the 0th class. We prove that formula (47) holds for all values of j by induction; assume that it holds for the class $(j-1)$, then substituting for $P_{...}$ and P_{jH-1} given by (47) in the equation (46) we obtain

$$P_{jH-R} = q^{-R+1}\left\{pP_0 q^{-H+1}\left(\sum_{v=0}^{j-1}(-1)^r p^{v-1}q^{vH-v}\left[\binom{(j-v)H+...-2}{v-1} + p\binom{(j-v)H+v-2}{v}\right]\right)\right.$$

$$- p^2 P_0 q^{-(j-1)H}\sum_{N=0}^{R}\left(\sum_{v=0}^{j-1}(-1)^v p^{v-1}q^{v(H-1)}\right.$$

$$\left.\times\left[\binom{(j-1-v)H+N+v-1}{v-1} + p\binom{(j-1-v)H+N+v-1}{v}\right]\right\}$$

$$= pP_0 q^{-H-R}\left\{1 + \sum_{r=1}^{j-1}(-1)^r p^{r-1}q^{vH-v}\left[\binom{(j-v)H+v-2}{v-1}\right.\right.$$

$$+ \sum_{N=0}^{R}\binom{(j-v)H+N+v-2}{v-2} + p\left[\binom{(j-v)H+v-2}{v} - \sum_{N=0}^{R}\binom{(j-v)H+N+v-2}{v-1}\right]$$

$$+ (-1)^j p^{j-1}q^{vH-1}\left[\sum_{N=0}^{R}\binom{N+j-2}{j-2} - p\sum_{N=0}^{R}\binom{N+j-2}{j-1}\right]\right\}.$$ (48)

Now in (48), consider

$$\left[\binom{(j-v)H+v-2}{v-1} + \sum_{N=0}^{R}\binom{(j-v)H+N+v-2}{v-2}\right];$$

since

$$\sum_{t=0}^{s}\binom{t}{w} = \sum_{t=w}^{s}\binom{t}{w} = \binom{n+1}{w+1},$$

then

$$\left[\binom{(j-v)H+v-2}{v-1} + \sum_{N=0}^{R}\binom{(j-v)H+N+v-2}{v-2}\right]$$

$$= \left[\binom{(j-v)H+v-2}{v-1} + \sum_{N=-(j-v)H}^{R}\binom{(j-v)H+N+v-2}{v-2} - \sum_{N=-(j-v)H}^{-1}\binom{(j-v)H+N+v-2}{v-2}\right]$$

$$= \binom{(j-v)H+R+v-1}{v-1}.$$

In exactly the same way, we reduce the other terms in (48), so that

$$\left\{\binom{(j-v)H+v-2}{v} + \sum_{N=0}^{R}\binom{(j-v)H+N+v-2}{v-1}\right\} = \binom{(j-v)H+R+v-1}{v},$$

and

$$\sum_{N=0}^{R}\binom{N+j-2}{j-2} + p\sum_{N=0}^{R}\binom{N+j-2}{j-1} = \binom{R+j-1}{j-1} + p\binom{R+j-1}{j}.$$

We thus obtain for P_{jH+R} in (48) the following expression:

$$P_{jH+R} = pP_0 q^{-(jH+R)}$$

$$\times \left\{1 + \sum_{v=1}^{j}(-1)^v p^{v-1}q^{v(H-1)}\left(\binom{(j-v)H+R+v-1}{v-1} + p\binom{(j-v)H+R+v-1}{v}\right)\right\}, \quad (49)$$

which is equivalent to (47). Now, since this has been shown to hold for the 0th class, then it will hold for the 1th class, and so on up to class (b); this proves that (47) (or the slightly different form (49)) is the solution required for the stationary distribution of $S(t)$ in the discrete case considered. We write out in full the formulae for a few classes so as to give an idea of their structure:

$$P_R = pP_0 q^{-R} \quad (R = 1, 2, ..., H-1);$$

$$P_{H+R} = pP_0 q^{-(H+R)}\left\{1 - q^{H-1}\left(1 + p\binom{R}{1}\right)\right\},$$

$$P_{2H+R} = pP_0 q^{-(2H+R)}\left\{1 - q^{H-1}\left(1 + p\binom{H+R}{1}\right) + pq^{2(H-1)}\left(\binom{R+1}{1} + p\binom{R+1}{2}\right)\right\}.$$

$$P_{3H+R} = pP_0 q^{-(3H+R)}\left\{1 - q^{H-1}\left(1 + p\binom{2H+R}{1}\right) + pq^{2(H-1)}\left(\binom{H+R+1}{1} + p\binom{H+R+1}{2}\right)\right.$$

$$\left. - p^2 q^{3(H-1)}\left(\binom{R+2}{2} + p\binom{R+2}{3}\right)\right\} \quad (R = 0, 1, ..., H-1).$$

For the complete solution in any particular case, the value of P_0 is also required; this can be found from the equation $\sum_{i=0}^{K-1} P_i = 1$.

We now proceed to the continuous case by letting $\delta t \to 0$, and consequently $\Delta = C\delta t \to 0$, where the constant C can now be interpreted as the rate of output per unit time. We allow Δ to tend to zero, and the previously defined integers K, H and R to tend to infinity in such a way that

(1) the maximum value of the storage function $K\Delta \to k$, or $(bH+U)\Delta \to bh+u$,

(2) the input $H\Delta \to h$,

(3) the values of the storage function $(jH+R)\Delta \to jh+r$, which we shall also write s $(j = 0, 1, ..., b; 0 \leqslant r < h)$.

As a result of this limiting process, the conditions of the problem which we have just considered for the storage function are changed; the function $S(t)$ has now a continuous range of values. $0 \leqslant S(t) < k$, where $k = bh + u$, and is defined for all time t (t continuous) in such a way that

(1) it is fed by a discrete input of Poisson type with parameter a, so that in a small interval of time δt there may be inputs of h or of no units with probabilities

$$p = \Pr\{I(t+\delta t) - I(t) = h\} = 1 - e^{-a\delta t} = 1 - e^{-(a/C)\Delta} = 1 - e^{-\mu\Delta},$$

$$q = \Pr\{I(t+\delta t) - I(t) = 0\} = e^{-a\delta t} = e^{-(a/C)\Delta} = e^{-\mu\Delta},$$

where $\mu = a/C$.

(2) the output is such that it has a constant rate C per unit time, provided the storage function is not zero, but is zero if the storage function is zero;

(3) the overflow is such that in any interval of time δt there is no overflow if

$$S(t) + I(t + \delta t) - I(t) \leqslant k,$$

but there is an overflow

$$S(t) + I(t + \delta t) - I(t) - k$$

if this is greater than zero.

We write the continuous stationary probability density for $S(t)$ as $f(s)$, so that

$$f(s)\,ds = \Pr\{s < S(t) < s + ds\}.$$

Now for $s = \lim_{\Delta \to 0} (jH + R)\Delta = jh + r$, we have that

$$f(s)\,ds = \lim_{\Delta \to 0} P_{jH+R}\Delta^{-1}ds;$$

this enables us to obtain from (45) and (49) the equations for $f(s)$ in various ranges of magnitude h. For $j = 0\ (0 < r < h)$ we have

$$f(r) = \lim_{\Delta \to 0} \Delta^{-1}pP_0 q^{-R} = \lim_{\Delta \to 0} P_0 \Delta^{-1}(1 - e^{-\mu\Delta})e^{\mu\Delta R}$$

$$= \mu e^{\mu r} P_0. \tag{50}$$

For $j = 1, 2, \ldots, b$ and $0 \leqslant r < h$, we have

$$f(jh+r) = \lim_{\Delta \to 0} \Delta^{-1}pP_0 q^{-(jH+R)}\left(1 + \sum_{v=1}^{j}(-1)^v p^{v-1}q^{v(H-1)}\right.$$

$$\left. \times \left\{\binom{(j-v)H + R + v - 1}{v-1} + p\binom{(j-v)H + R + v - 1}{v}\right\}\right)$$

$$= \lim_{\Delta \to 0} \Delta^{-1}(1 - e^{-\mu\Delta})e^{\mu\Delta(jH+R)}P_0\left(1 + \sum_{v=1}^{j}(-1)^v e^{-\mu v\Delta(H-1)}\right.$$

$$\left. \times \left\{(\mu\Delta)^{v-1}\binom{(j-v)H + R + v - 1}{v-1} + (\mu\Delta)^v\binom{(j-v)H + R + v - 1}{v}\right\}\right)$$

$$= \mu e^{\mu(jh+r)}P_0\left(1 + \sum_{v=1}^{j}(-1)^v e^{-\mu rh}\left\{\mu^{v-1}\frac{((j-v)h+r)^{v-1}}{(v-1)!} + \frac{\mu^v((j-v)h+r)^v}{v!}\right\}\right). \tag{51}$$

In addition to these probability densities, which are continuous in the intervals in which they are defined, there is also a concentration of probability P_0 at $s = 0$. To complete the solution, we require the value of P_0; this will be given in any particular case where the value of $k = bh + u$ is known by

$$P_0 + \sum_{j=0}^{b-1}\int_{r=0}^{h} f(jh+r)\,dr + \int_{r=0}^{u} f(bh+r)\,dr = 1. \tag{52}$$

We write out in full some of the formulae for the $f(jh+r)$, in order to indicate their structure:

$$
\left.
\begin{aligned}
f(r) &= \mu e^{\mu r} P_0 \quad (0 < r < h), \\
f(h+r) &= \mu e^{\mu(h+r)}P_0\{1 - e^{-\mu h}(1 + \mu r)\}, \\
f(2h+r) &= \mu e^{\mu(2h+r)}P_0\left\{1 - e^{-\mu h}(1 + \mu(h+r)) + \mu e^{-2\mu h}\left(r + \mu\frac{r^2}{2!}\right)\right\}, \\
f(3h+r) &= \mu e^{\mu(3h+r)}P_0\left\{1 - e^{-\mu h}(1 + \mu(2h+r)) + \mu e^{-2\mu h}\left((h+r) + \mu\frac{(h+r)^2}{2!}\right)\right. \\
&\qquad\qquad \left. - \mu^2 e^{-3\mu h}\left(\frac{r^2}{2!} + \mu\frac{r^3}{3!}\right)\right\} \quad (0 \leqslant r < h).
\end{aligned}
\right\}
\tag{53}
$$

It is interesting to note that these formulae for $f(jh+r)$ $(j = 1, 2, ..., b)$ can also be obtained from the limit of equation (46), when $\Delta \to 0$,

$$f(jh+r) = \lim_{\Delta \to 0} P_{jH+R} \Delta^{-1}$$
$$= \lim_{\Delta \to 0} q^{-(R+1)} \left\{ P_{jH-1} \Delta^{-1} - p\Delta^{-1} \sum_{N=0}^{R} q^N (P_{(j-1)H+N} \Delta^{-1}) \Delta \right\}. \tag{54}$$

Now for $j = 1$, we have

$$f(h+r) = \lim_{\Delta \to 0} P_{H+R} \Delta^{-1}$$
$$= \lim_{\Delta \to 0} q^{-(R+1)} \left\{ P_{H-1} \Delta^{-1} - p\Delta^{-1} \sum_{N=1}^{R} q^N (P_N \Delta^{-1}) \Delta - p\Delta^{-1} P_0 \right\}$$
$$= e^{\mu r} f(h-0) - \mu e^{\mu r} \left\{ \int_0^r e^{-\mu n} f(n)\, dn + P_0 \right\}, \tag{55}$$

since $P_N \Delta^{-1} \to f(n)$, where $N\Delta \to n$ for all $N \neq 0$, but there is a concentration of probability P_0 at $s = 0$. For $j = 2, 3, ..., b$ the equation (54) gives

$$f(jh+r) = e^{\mu r} f(jh-0) - \mu e^{\mu r} \int_0^r e^{-\mu n} f((j-1)h+n)\, dn. \tag{56}$$

The formulae (53) can be obtained by putting the formula (50) for $f(r)$ in (55) and integrating it to find $f(h+r)$, then repeating this process for $f(jh+r)$ by using (56). More generally, it can be proved by induction that our solutions (50) and (51) satisfy the integral equations (55) and (56).

It is found that (52) does not lend itself to the computation of P_0 as well as an alternative equation for the integral of the probabilities, which we now obtain. To do so, we write the probability density $f(s)$ for the various intervals in the following manner:

$$f(s) = \mu e^{\mu s} P_0 \quad (0 < s < h);$$

$$f(s) = \mu e^{\mu s} P_0 \left(1 + \sum_{v=1}^{j} (-1)^v e^{-\mu v h} \left\{ \mu^{v-1} \frac{(s-vh)^{v-1}}{(v-1)!} + \mu^v \frac{(s-vh)^v}{v!} \right\} \right)$$

$$(jh \leqslant s < (j+1)h;\ j = 1, 2, ..., b).$$

The equation (52) can then be written as

$$P_0 \left(1 + \int_0^{bh+u} \mu e^{\mu s}\, ds + \sum_{j=1}^{b} (-1)^j \int_{s-jh}^{bh+u} \mu e^{\mu(s-jh)} \left\{ \mu^{j-1} \frac{(s-jh)^{j-1}}{(j-1)!} + \mu^j \frac{(s-jh)^j}{j!} \right\} ds \right) = 1,$$

which shortens considerably the work involved in the computation of P_0.

I am greatly indebted to Prof. P. A. P. Moran for suggesting several of the problems, and for his extremely helpful discussion and criticism of the work throughout all stages.

REFERENCES

FELLER, W. (1950). *An Introduction to Probability Theory and its Applications.* New York: John Wiley and Sons.

FRAZER, R. A., DUNCAN, W. J. & COLLAR, A. R. (1947). *Elementary Matrices.* Cambridge University Press.

MORAN, P. A. P. (1954). A probability theory of dams and storage systems. *Aust. J. Appl. Sci.* 5, no. 2, pp. 116–24.

MORAN, P. A. P. (1955). A probability theory of dams and storage systems. II (to appear).

PITT, H. R. (1946). A theorem on random functions with applications to a theory of provisioning. *J. Lond. Math. Soc.* 21, 16–22.

APPROXIMATE CONFIDENCE INTERVALS

III. A BIAS CORRECTION

By M. S. BARTLETT
University of Manchester

In my second paper on approximate confidence intervals (Bartlett, 1953, to be referred to as II), where a general method of eliminating 'nuisance parameters' was discussed, it was claimed (see top of p. 310) that the effect of substituting maximum-likelihood estimates for the nuisance parameters could be neglected to the first order ($O(1/\sqrt{n})$) of approximation. I regret to say that this assertion is not in general correct, and the statistic T, obtained after such substitution may have a bias of $O(1/\sqrt{n})$ which requires correction. It is even necessary in the expansion

$$T_e = T + (\hat{\theta}_2 - \theta_2)\frac{\partial T}{\partial \theta_2} + \tfrac{1}{2}(\hat{\theta}_2 - \theta_2)^2\frac{\partial^2 T}{\partial \theta_2^2} + \cdots \tag{1}$$

to make allowance not only for the second term, but also for the *third* term, on the right-hand side of (1), or to use the equivalent approximation to this order given in equation (14) of II. For convenience we may write (cf. equation (11) of II)

$$T \equiv a\frac{\partial L}{\partial \theta_1} + b\frac{\partial L}{\partial \theta_2}. \tag{2}$$

where $a \equiv 1/\sqrt{I_{11.2}}$, $b \equiv -I_{12}\,(I_{22}\sqrt{I_{11.2}})$. Then, noting that $\partial T/\partial \theta_2$ is $O(1)$, and $\partial^2 T/\partial \theta_2^2$ $O(\sqrt{n})$, we find, omitting any terms of smaller order of magnitude than $1/\sqrt{n}$,

$$T_e - T = \frac{1}{I_{22}}\frac{\partial L}{\partial \theta_2}\left[a\frac{\partial^2 L}{\partial \theta_1\,\partial \theta_2} + b\frac{\partial^2 L}{\partial \theta_2^2} + \frac{\partial L}{\partial \theta_1}\frac{\partial a}{\partial \theta_2} + \frac{\partial L}{\partial \theta_2}\frac{\partial b}{\partial \theta_2}\right]$$
$$+ \frac{1}{2I_{22}^2}\left(\frac{\partial L}{\partial \theta_2}\right)^2\left[aE\left\{\frac{\partial^3 L}{\partial \theta_1\,\partial \theta_2^2}\right\} + bE\left\{\frac{\partial^3 L}{\partial \theta_2^3}\right\} - 2I_{12}\frac{\partial a}{\partial \theta_2} - 2I_{22}\frac{\partial b}{\partial \theta_2}\right]. \tag{3}$$

Hence, making use of relations such as

$$E\left\{\frac{\partial L}{\partial \theta_2}\frac{\partial^2 L}{\partial \theta_1\,\partial \theta_2}\right\} = -E\left\{\frac{\partial^3 L}{\partial \theta_1\,\partial \theta_2^2}\right\} - \frac{\partial I_{12}}{\partial \theta_2}, \tag{4}$$

we obtain from (3)

$$E\{T_e\} = -\frac{a}{2I_{22}}\left[E\left\{\frac{\partial^3 L}{\partial \theta_1\,\partial \theta_2^2}\right\} + 2\frac{\partial I_{12}}{\partial \theta_2}\right]$$
$$- \frac{b}{2I_{22}}\left[E\left\{\frac{\partial^3 L}{\partial \theta_2^3}\right\} + 2\frac{\partial I_{22}}{\partial \theta_2}\right] + O\left(\frac{1}{n}\right). \tag{5}$$

When $I_{12} = 0$, this reduces to

$$E\{T_e\} \sim -\tfrac{1}{2}E\left\{\frac{\partial^3 L}{\partial \theta_1\,\partial \theta_2^2}\right\}\bigg/(I_{22}\sqrt{I_{11}}). \tag{6}$$

For reference, the generalizations of (3) and (5) in the case of several nuisance parameters $\theta_i\ (i = 2, \ldots, m)$ are

$$T_e - T \sim I^{ik}\frac{\partial L}{\partial \theta_i}\left[A\frac{\partial^2 L}{\partial \theta_1\,\partial \theta_k} + B_j\frac{\partial^2 L}{\partial \theta_j\,\partial \theta_k} + \frac{\partial L}{\partial \theta_1}\frac{\partial A}{\partial \theta_k} + \frac{\partial L}{\partial \theta_j}\frac{\partial B_j}{\partial \theta_k}\right]$$
$$+ \tfrac{1}{2}I^{ik}\frac{\partial L}{\partial \theta_i}\frac{\partial L}{\partial \theta_g}I^{gh}\left[AE\left\{\frac{\partial^3 L}{\partial \theta_1\,\partial \theta_k\,\partial \theta_h}\right\} + B_jE\left\{\frac{\partial^3 L}{\partial \theta_j\,\partial \theta_k\,\partial \theta_h}\right\} - 2I_{1k}\frac{\partial A}{\partial \theta_h} - 2I_{jk}\frac{\partial B_j}{\partial \theta_h}\right], \tag{3a}$$

where
$$T \equiv A \frac{\partial L}{\partial \theta_1} + B_j \frac{\partial L}{\partial \theta_j},$$

say, and

$$E\{T_c\} \sim -\tfrac{1}{2}A\left[E\left\{\frac{\partial^3 L}{\partial\theta_1\,\partial\theta_k\,\partial\theta_h}\right\} I^{kh} + 2I^{ik}\frac{\partial I_{1k}}{\partial\theta_i}\right] - \tfrac{1}{2}B_j\left[E\left\{\frac{\partial^3 L}{\partial\theta_j\,\partial\theta_k\,\partial\theta_h}\right\} I^{kh} + 2I^{ik}\frac{\partial I_{jk}}{\partial\theta_i}\right], \quad (5a)$$

the summation convention being understood for g, h, i, j, k $(2, \dots, m)$, and I^{ij} being the inverse of I_{ij}.

Provided T_c is corrected for its bias, it is found further from the formulae (3) and (3a) that its second and third moments are then the same as those of T to $O(1/\sqrt{n})$.

Examples. (i) and (ii). The two examples discussed in §§ 6 and 7 of II are not affected by this correction. In the analysis of variance problem of § 6, $I_{12} \neq 0$, but it will be found that $E(T_c) \sim 0$ from (5) above. (I have learnt that more direct investigations of the asymptotic confidence interval for this problem, using the methods first developed by Dr B. L. Welch in the Behrens–Fisher problem, have recently[*] been made by J. R. Green and by A. Huitson; and, in particular, Dr Huitson has checked the consistency of my solution with his own results.)

Similarly, no bias correction arises in the time-series problem of § 7, though there is unfortunately a further misstatement in equation (23), which should be corrected. The maximum-likelihood estimate of $\sigma_1^2 \equiv \alpha$, say, should have been given as

$$\hat{\alpha} = \sum_{r=1}^{n} (X_r - \beta X_{r-1})^2/n. \quad (7)$$

The confidence interval from (24) then becomes

$$\beta = \hat{\beta} \pm \frac{\mu}{\sqrt{n}} \sqrt{(1-\hat{\beta}^2)} + \frac{\hat{\beta}}{n} + O\left(\frac{1}{n^{\frac{3}{2}}}\right) \quad (8)$$

(cf. my contribution to the discussion at the Royal Statistical Society Symposium on Interval Estimation, 12 May 1954).

(iii) To illustrate the use of the bias correction where it is not zero, consider the last example above when the mean value m of X is not zero and requires estimation. The log likelihood function (apart from end-corrections of relative order $1/n$) is now

$$L = -\frac{n}{2}\log\alpha - \frac{1}{2}\sum_{r=1}^{n} Y_r^2/\alpha, \quad (9)$$

where
$$Y_r = (X_r - m) - \beta(X_{r-1} - m). \quad (10)$$

The L derivatives with respect to α and β are as before, with $\xi_r = X_r - m$ in place of X_r. Also

$$\frac{\partial L}{\partial m} = \frac{(1-\beta)}{\alpha}\sum_{r=1}^{n}(\xi_r - \beta\xi_{r-1}),$$

$$\hat{m} \sim \sum_{r=1}^{n} X_r/n,$$

$$I_{mm} = (1-\beta)^2 n/\alpha, \quad I_{\alpha m} = 0, \quad I_{\beta m} = 0,$$

$$\frac{\partial^3 L}{\partial\beta\,\partial m^2} = 2n(1-\beta)/\alpha = E\left\{\frac{\partial^3 L}{\partial\beta\,\partial m^2}\right\}.$$

Hence the quantity

$$T_c \equiv \frac{\sum_{r=1}^{n}(X'_r - \beta X'_{r-1}) X'_{r-1}/\hat{\alpha}'}{\sqrt{\{n/(1-\beta^2)\}}}, \quad (11)$$

[*] See references.

where $X'_r \equiv X_r - \hat{m}$, and $\hat{\alpha}'$ is like $\hat{\alpha}$ in (7) but with X'_r in place of X_r, has bias

$$E\{T_e\} \sim -\frac{1}{(1-\beta)\sqrt{I_{\beta\beta}}},$$

and the equation for β becomes

$$T_e + \frac{\sqrt{(1-\beta^2)}}{(1-\beta)\sqrt{n}} - \frac{\beta(\mu^2-1)}{\sqrt{(1-\beta^2)n}} = \dot{} \mu. \tag{12}$$

or

$$\beta - \hat{\beta} \doteq \frac{\mu}{\sqrt{n}}\sqrt{(1-\hat{\beta}^2)} + \frac{1+2\hat{\beta}}{n} + O\left(\frac{1}{n^{\frac{1}{2}}}\right) \tag{13}$$

in place of (8), where now

$$\hat{\beta} \equiv \sum_{r=1}^{n} X'_r X'_{r-1} \Big/ \sum_{r=1}^{n} (X'_{r-1})^2.$$

(iv) Finally, in view of the above amendments, it seemed advisable to check these results to $O(1/\sqrt{n})$ by means of a fairly complicated example for which the result is already known. A useful case is that of the classical correlation coefficient ρ, with the two standard deviations α and β as nuisance parameters. We take, apart from an additive constant,

$$L = -n \log \{\alpha\beta\sqrt{(1-\rho^2)}\} - \frac{n}{2(1-\rho^2)}\left[\frac{s_1^2}{\alpha^2} - \frac{2\rho r s_1 s_2}{\alpha\beta} + \frac{s_2^2}{\beta^2}\right]. \tag{14}$$

where r, s_1, s_2 are the usual sample estimates of ρ, α and β respectively (the means for convenience are already eliminated). Then (cf. M. G. Kendall's *Advanced Theory of Statistics*, vol. 2, ex. 17.18)

$$I_{\alpha\alpha} = \frac{n(2-\rho^2)}{\alpha^2(1-\rho^2)}, \quad I_{\beta\beta} = \frac{n(2-\rho^2)}{\beta^2(1-\rho^2)}, \quad I_{\rho\rho} = \frac{n(1+\rho^2)}{(1-\rho^2)^2},$$

$$I_{\alpha\beta} = \frac{-n\rho^2}{\alpha\beta(1-\rho^2)}, \quad I_{\alpha\rho} = \frac{-n\rho}{\alpha(1-\rho^2)}, \quad I_{\beta\rho} = \frac{-n\rho}{\beta(1-\rho^2)}. \tag{15}$$

It is found further from (14) that

$$E\left\{\frac{\partial^3 L}{\partial\alpha^3}\right\} = \frac{2n(5-2\rho^2)}{\alpha^3(1-\rho^2)}, \qquad E\left\{\frac{\partial^3 L}{\partial\alpha^2\partial\beta}\right\} = \frac{-2n\rho^2}{\alpha^2\beta(1-\rho^2)},$$

$$E\left\{\frac{\partial^3 L}{\partial\alpha^2\partial\rho}\right\} = \frac{-2n\rho(1-\rho^2)}{\alpha^2(1-\rho^2)^2}, \qquad E\left\{\frac{\partial^3 L}{\partial\alpha\partial\beta\partial\rho}\right\} = \frac{n\rho(1+\rho^2)}{\alpha\beta(1-\rho^2)^2}, \tag{16}$$

$$E\left\{\frac{\partial^3 L}{\partial\alpha\partial\rho^2}\right\} = \frac{2n(1+\rho^2)}{\alpha(1-\rho^2)^2}, \qquad E\left\{\frac{\partial^3 L}{\partial\rho^3}\right\} = \frac{-4n\rho(3+\rho^2)}{(1-\rho^2)^3},$$

the expressions for the remaining quantities with α and β interchanged being obvious from symmetry.

The statistic T uncorrelated with $\partial L/\partial\alpha$ and $\partial L/\partial\beta$ is

$$T \equiv \left(\frac{\partial L}{\partial\rho} + \frac{\alpha\rho}{2(1-\rho^2)}\frac{\partial L}{\partial\alpha} + \frac{\beta\rho}{2(1-\rho^2)}\frac{\partial L}{\partial\beta}\right)\Big/\sqrt{I_{\rho\rho.\alpha\beta}}. \tag{17}$$

where $I_{\rho\rho.\alpha\beta} = n/(1-\rho^2)^2$. From the above formulae, and the general expression for $E\left\{\frac{\partial L}{\partial\theta_i}\frac{\partial L}{\partial\theta_j}\frac{\partial L}{\partial\theta_k}\right\}$ in terms of $E\left\{\frac{\partial^3 L}{\partial\theta_i\partial\theta_j\partial\theta_k}\right\}$ and derivatives of I_{ij} given in II, it is found (after some algebra) that the skewness of T in (17) is zero to $O(1/\sqrt{n})$.

We now replace T by T_e, the maximum-likelihood estimates of α and β (ρ known) being

$$\hat{\alpha} = s_1\sqrt{\left(\frac{1-\rho r}{1-\rho^2}\right)}, \quad \hat{\beta} = s_2\sqrt{\left(\frac{1-\rho r}{1-\rho^2}\right)}. \tag{18}$$

T in (17) then reduces to

$$T_e \equiv \frac{\sqrt{n}}{1-\rho^2}\left(\frac{\partial L}{\partial \rho}\right)_e = \frac{\sqrt{n}(r-\rho)}{1-\rho r}; \tag{19}$$

the functional form for r in (19) should be noted. From the general bias formula (5a), we find (noting that $I^{\alpha\alpha} = \frac{1}{4}(2-\rho^2)\,\alpha^2, n, \ I^{\alpha\beta} = \frac{1}{4}\rho^2\alpha\beta/n$)

$$E(T_e) = \frac{1}{2}\rho/\sqrt{n}. \tag{20}$$

Hence our confidence interval is given by

$$\frac{\sqrt{n}(r-\rho)}{1-\rho r} - \frac{\rho}{2\sqrt{n}} = \pm\frac{\mu}{\sqrt{n}}, \tag{21}$$

which agrees with the known results

$$E(z) \sim \zeta + \tfrac{1}{2}\rho/n, \quad \sigma^2(z) \sim 1/n, \quad \gamma_1(z) \sim 0,$$

where

$$z = \tfrac{1}{2}\log_e\frac{1+r}{1-r}, \quad \zeta = \tfrac{1}{2}\log_e\frac{1+\rho}{1-\rho}.$$

While the above method is not of course required for this example, it is of some interest that the result (21) has been obtained merely by straightforward differentiation of the log likelihood function (14). It is, however, apparent that the algebra involved in getting the next term in the expansion becomes in general so intractable that a more direct attack on individual problems is then usually the more promising.

REFERENCES

BARTLETT, M. S. (1953). Approximate confidence intervals. II. More than one unknown parameter. *Biometrika*, **40**, 306.

GREEN, J. R. (1954). A confidence interval for variance components. *Ann. Math. Statist.* 25. 671.

HUTSON, A. (1954). Ph.D. Thesis, University of Leeds.

THE THEORY OF CORRELATION BETWEEN TWO CONTINUOUS VARIABLES WHEN ONE IS DICHOTOMIZED

By ROBERT F. TATE

University of Washington†

1. INTRODUCTION AND SUMMARY

The problem of biserial correlation arises when one is sampling from a bivariate normal population in which one of the variables has been dichotomized, giving rise to only two observable values, say 0 and 1, and one wishes to use this dichotomized sample to estimate, or to test hypotheses concerning, the correlation coefficient ρ of the original bivariate normal distribution. The problem of biserial correlation occurs frequently in psychological work, especially in test construction and validation.

The term biserial correlation was introduced by Karl Pearson (1909), who was the first to perceive the statistical importance of this particular type of problem. He proposed as an estimator the sample biserial correlation coefficient. The asymptotic variance of this estimator was derived by Soper (1913). Much literature exists on the subject of how best to compute Pearson's coefficient. In this connexion the reader should see Du Bois (1942), Dunlap (1936) and Royer (1941).

Prof. Harold Hotelling realized some years ago that the existing methods for dealing with the problem of biserial correlation were far from satisfactory, and suggested to the author that the whole situation be reconsidered. The results of this examination are contained in the present paper.

§ 2 contains a list of most of the notation which has been adopted, and § 3 deals with the mathematical model. In § 4 the question of maximum likelihood is treated. Asymptotic variances are derived for the estimators $\hat{\omega}$ and $\hat{\rho}$. The asymptotic variance for $\hat{\rho}$ is compared with the approximate expression arrived at by Maritz (1953) when he considered a somewhat restricted model. Both expressions are shown to achieve their minimum value at $\omega = 0$ when ρ is fixed.

Matters concerning asymptotic normality and asymptotic efficiency are also considered.

An appraisal of r^*, the sample biserial correlation coefficient, is given in detail in § 5. It is shown to have asymptotic efficiency for estimating ρ which is 1 when $\rho = 0$, but which approaches 0 when $|\rho|$ approaches 1. The well-known fact that r^* may be greater than 1 is pointed out and some notion of the magnitude of r^* is obtained by a consideration of the product-moment correlation coefficient r. Asymptotic normality of r^* is verified by the use of a theorem of Cramér. The asymptotic standard deviation is tabulated at the end of the paper (Table 2). A proof is given for the customarily assumed fact that the asymptotic variance has a minimum for fixed ρ when $\omega = 0$. For the case $\omega = 0$ an approximate variance stabilizing transformation is derived. Calculations pertaining to this transformation may

be carried out by using a table (see Fisher, 1946, Table V B) for the function $\tanh^{-1} r$. This result should prove useful in many situations.

§ 6 is devoted to a discussion of an iterative method of solution for the likelihood equations. The method is essentially Newton's method for two variables, the calculated values ω^*, r^* being used to start the iteration. The computations are not really prohibitive, considering the importance of the problem, and are to a certain extent organizable for punched-cards methods. An example is given with all of the calculations illustrated. Values of $\phi(x)$, the reciprocal of Mills's ratio, are required for the solution of the likelihood equations. These may be obtained from the tables published as a separate contribution, immediately following the present paper.

Two matters of some importance which are not considered in the present paper are:

(1) An investigation of the bias of $\hat{\rho}$.
(2) The numerical tabulation of the asymptotic variances of $\hat{\omega}$ and $\hat{\rho}$.

Further study is indicated on at least the second point.

2. NOTATION

To eliminate the distraction of searching through the text, we shall list here most of the symbols and notational devices used:

$$\psi(x,y) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right], \text{ the bivariate normal density.}$$

$$\lambda(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}, \text{ the normal density.}\dagger$$

$$p(x) = \int_x^{+\infty} \lambda(t)\,dt, \quad q(x) = 1 - p(x).\dagger$$

$$\phi(x) = \frac{\lambda(x)}{p(x)}, \text{ the reciprocal of Mills's ratio.}\dagger$$

$$\xi(x,\omega) = \int_\omega^{+\infty} \psi(x,y)\,dy.$$

$$\eta(x,\omega) = \int_{-\infty}^{\omega} \psi(x,y)\,dy.$$

X	the undichotomized normal random variable.
Y	the dichotomized normal random variable.
ω	the point of dichotomy of Y, measured in standard units.
Z	the discrete random variable induced by the dichotomization of Y.
$f(x,z)$	the joint density of the random variables X and Z.
r^*	the sample biserial correlation coefficient.
$AV(r^*)$	the asymptotic variance of r^*.
$AEff(r^*)$	the asymptotic efficiency of r^* for estimating ρ.
$\mathcal{N}(\mu, \sigma^2)$	a normal random variable with mean μ and variance σ^2.

† [*Editorial Note.* To bring this notation into conformity with that of the tables printed on pp. 217–221 below and with that used in the recently published *Biometrika Tables for Statisticians*, vol. 1, it is necessary to write $Z = \lambda(x)$, $Q = p(x)$, $P = 1 - p(x)$, so that $\phi(x) = Z/Q$ or Z/P according as x is > 0 or < 0.]

3. MATHEMATICAL MODEL

Let (X, Y) have the bivariate normal distribution $\psi\{(x-\mu)/\sigma, (y-\tau)/\tau\}$. Let Z be a dichotomy of Y, with the point of dichotomy ω measured in standard units. Without losing any generality we may set $\nu = 0$ and $\tau = 1$. Z is thus a random variable which takes the value 1 when $Y \geqslant \omega$ and the value 0 when $Y < \omega$. Obviously,

$$P(Z=1) = \int_{\omega}^{+\infty} \lambda(y)\, dy = p(\omega), \quad P(Z=0) = q(\omega). \tag{3.1}$$

Consider a sample of n independent random vectors $(X_1, Z_1), (X_2, Z_2), \ldots, (X_n, Z_n)$. The problem of biserial correlation consists of finding a suitable function of (X_i, Z_i) $(i = 1, 2, \ldots, n)$ with which to estimate ρ.

Karl Pearson (1909) introduced the estimator r^* ('biserial r'), which we express in the following form:[†]

$$r^* = \frac{\frac{1}{n}\Sigma(X_i - \bar{X})(Z_i - \bar{Z})}{\left\{\frac{1}{n}\Sigma(X_i - \bar{X})^2\right\}^{\frac{1}{2}} \lambda(T)} = r\frac{\left\{\frac{1}{n}\Sigma(Z_i - \bar{Z})^2\right\}^{\frac{1}{2}}}{\lambda(T)}, \tag{3.2}$$

where r is the product-moment correlation coefficient of (X_i, Z_i), and T is the solution of the equation[‡]

$$\int_{T}^{+\infty} \lambda(y)\, dy = \bar{Z}. \tag{3.3}$$

r^* will be discussed completely in §5. For the present we shall merely state the asymptotic variance obtained by Soper (1913):

$$AV(r^*) = \frac{1}{n}\left\{\rho^4 + \rho^2\left[\frac{pq\omega^2}{\lambda^2} + \frac{(2p-1)\omega}{\lambda} - \frac{5}{2}\right] + \frac{pq}{\lambda^2}\right\}, \tag{3.4}$$

where the functions p, q and λ all have argument ω. $\sqrt{\{Av(r^*)\}}$ is given in Table 2 at the end of the paper. In view of symmetry about the values $\rho = 0$ and $p(\omega) = \frac{1}{2}$, the tabulation is given for $\rho = 0$ to 1 in steps of 0.10, and for $p = 0.05$ to 0.50 in steps of 0.05.

Since the random variable Z takes the value 0 or 1, the joint density of (X, Z) can be written

$$f(x, z) = zf(x, 1) + (1-z)f(x, 0), \tag{3.5}$$

where

$$f(x, 0) = \int_{-\infty}^{\omega} \Psi(x, y)\, dy, \quad f(x, 1) = \int_{\omega}^{+\infty} \Psi(x, y)\, dy, \tag{3.6}$$

with $\Psi(x, y)$ denoting the bivariate normal density, $\psi\{(x-\mu)/\sigma, y\}$, with means μ and 0, variances σ^2 and 1, and correlation ρ. §§4 and 6 are devoted to a discussion of the likelihood function $\Pi f(x_i, z_i)$.

4. PROPERTIES OF THE MAXIMUM-LIKELIHOOD ESTIMATORS

As the likelihood function stands it may be expressed as

$$L(\mu, \sigma^2, \omega, \rho) = \Pi\left\{z_i \xi\left(\frac{x_i - \mu}{\sigma}, \omega\right) + (1 - z_i)\eta\left(\frac{x_i - \mu}{\sigma}, \omega\right)\right\}. \tag{4.1}$$

Maritz (1953) considered the restricted model with $\mu = 0$, $\sigma^2 = 1$. Using biserial data (X_i, Z_i) $(i = 1, 2, \ldots, n)$, he introduces a grouping of the X observations, and then considers

[†] All Σ and Π symbols with index i will have limits 1 to n.
[‡] The reason for this definition of T will be apparent in §5, when we show that r^* is consistent and asymptotically normal.

the observations to be concentrated at their respective cell mid-points. This leads to a neat solution of the problem by probit-analysis methods. A proof of the convergence of this method as the grouping becomes finer must depend on a close examination of the limiting processes involved. Specifically, it is necessary that as the cell width becomes small, and the sample size large, each cell must contain sufficiently many observations that the ratio of the number of X observations whose corresponding Z observations are 1 to the number of X observations provides a valid approximation to the conditional probability that $Z = 1$. Instead of attempting a discussion of this point, we shall derive the asymptotic variances for the four-parameter problem (and, as a by-product, for the two-parameter problem), and in § 6 discuss an iterative method for obtaining $\hat{\omega}$ and $\hat{\rho}$. This method, while more time-consuming than that of Maritz, does not require grouping. It should be noted in this con-nexion that Tocher's exact method (see Tocher, 1949, pp. 9–11), also known as the 'scoring' method, does not help in this case, owing to the difficulty of obtaining expected second partial derivatives of L.

The likelihood situation of Maritz differs from ours because of the fact that when μ and σ^2 are set equal to 0 and 1 respectively, the customary method for obtaining asymptotic variances of $\hat{\omega}$ and $\hat{\rho}$ by an inversion of the information matrix leads to smaller variances. As far as the solution for $\hat{\omega}$ and $\hat{\rho}$ is concerned, the results will remain the same after a slight transformation.†

It may be remarked without dwelling at any length on regularity conditions that those given by Cramér (1946, p. 500) may be easily verified, since $f(x, 0)$ and $f(x, 1)$ are both in-tegrals of bivariate normal densities. Consequently, $\hat{\mu}, \hat{\sigma}^2, \hat{\omega}$ and $\hat{\rho}$ will be asymptotically normal and asymptotically efficient estimators of the corresponding parameters. We now use the information matrix technique to find the asymptotic variances of $\hat{\omega}$ and $\hat{\rho}$.

Theorem I. *The asymptotic variances of $\hat{\omega}$ and $\hat{\rho}$ are*

$$AV(\hat{\omega}) = \frac{1-\rho^2}{n} \left\{ \frac{\int_{-\infty}^{+\infty} (x-\rho\omega)^2 g \, dx}{\int_{-\infty}^{+\infty} g \, dx \int_{-\infty}^{+\infty} x^2 g \, dx - \left(\int_{-\infty}^{+\infty} xg \, dx \right)^2} \right\} + \frac{\rho^2(\rho^2\omega^2 + 2)}{n},$$

$$AV(\hat{\rho}) = \frac{(1-\rho^2)^3}{n} \left\{ \frac{\int_{-\infty}^{+\infty} g \, dx}{\int_{-\infty}^{+\infty} g \, dx \int_{-\infty}^{+\infty} x^2 g \, dx - \left(\int_{-\infty}^{+\infty} xg \, dx \right)^2} \right\} + \frac{\rho^2(1-\rho^2)^2}{n},$$

where

$$g(x, \omega, \rho) = \lambda(x) \phi \left(\frac{\omega - \rho x}{\sqrt{(1-\rho^2)}} \right) \phi \left(-\frac{\omega - \rho x}{\sqrt{(1-\rho^2)}} \right).$$

Proof. Using expression (4·1) for L, and letting δ^2 refer to any of the ten-second order partial operators, we obtain the fundamental relation

$$E(\delta^2 \log L) = nqE_0(\delta^2 \log \eta) + npE_1(\delta^2 \log \xi). \tag{4·2}$$

where E_0 and E_1 mean conditional expectation with respect to the conditional densities of X given $Y < \omega$ and $Y \geqslant \omega$ respectively.

---

† The author expresses his indebtedness to the referee for this fact, and for its proof which will be given later.

For each of the possible operators δ^2 the calculation of (4·2) proceeds in about the same way. We compute, as an illustration,

$$E\left(\frac{\partial^2 \log L}{\partial \omega^2}\right).$$

It may be shown that

$$\Psi(x \mid Y = \omega) = \frac{1}{q}\eta\left(\frac{x-\mu}{\sigma}, \omega\right), \quad \Psi(x \mid Y = \omega) = \frac{1}{p}\xi\left(\frac{x-\mu}{\sigma}, \omega\right).$$

After some differentiation we get

$$nqE_0\left(\frac{\partial^2 \log \eta}{\partial \omega^2}\right) = -n\int_{-\infty}^{+\infty}\frac{\Psi^2(x,\omega)}{\eta\{(x-\mu)/\sigma, \omega\}}\,dx. \left.\begin{array}{c} \\ \\ \end{array}\right|$$

$$npE_1\left(\frac{\partial^2 \log \xi}{\partial \omega^2}\right) = -n\int_{-\infty}^{+\infty}\frac{\Psi^2(x,\omega)}{\xi\{(x-\mu)/\sigma, \omega\}}\,dx. \qquad (4\cdot3)$$

Combining the terms of (4·3) according to (4·2), making use of the relation

$$\int_{-\infty}^{+\infty}\Psi^2(x,\omega)\left\{\frac{1}{\eta\left(\frac{x-\mu}{\sigma}, \omega\right)} + \frac{1}{\xi\left(\frac{x-\mu}{\sigma}, \omega\right)}\right\}\,dx = \int_{-\infty}^{+\infty}g(x,\omega,\rho)\,dx,$$

and performing similar computations for the expectations of the other second partials of $\log L$, we arrive at the information matrix†

$$\begin{bmatrix} -\dfrac{a_0}{1-\rho^2} & \dfrac{a_1-\rho\omega a_0}{(1-\rho^2)^2} & -\dfrac{\rho a_0}{\sigma(1-\rho^2)} & -\dfrac{\rho a_1}{\sigma(1-\rho^2)} \\[2ex] & -\dfrac{a_2-2\rho\omega a_1+\rho^2\omega^2 a_0}{(1-\rho^2)^3} & \dfrac{a_1\rho-\rho^2\omega a_0}{\sigma(1-\rho^2)^2} & \dfrac{a_2\rho-\rho^2\omega a_1}{\sigma(1-\rho^2)^2} \\[2ex] \text{symmetric} & & -\dfrac{1-\rho^2+\rho^2 a_0}{\sigma^2(1-\rho^2)} & -\dfrac{\rho^2 a_1}{\sigma^2(1-\rho^2)} \\[2ex] & & & -\dfrac{2(1-\rho^2)+\rho^2 a_2}{\sigma^2(1-\rho^2)} \end{bmatrix},$$

where

$$a_k = \int_{-\infty}^{+\infty}x^k g(x,\omega,\rho)\,dx. \qquad (4\cdot4)$$

The asymptotic variances of $\hat{\omega}$ and $\hat{\rho}$, obtained after inversion, correspond with the expressions of the theorem.

The two-parameter problem of Maritz, solved by considering the upper left 2×2 submatrix, yields asymptotic variances which are the leading terms of the expressions given in the theorem; they coincide with his results.

The role played by ω in $AV(\hat{\rho})$ is partially described by the following theorem:

THEOREM II. $AV(\hat{\rho})$ is a minimum for each ρ when $\omega = 0$.

Proof. For the case $\rho = 0$ the proof follows from Theorem IV of the next section. For the case $\rho \neq 0$, make the transformation $(\omega - \rho x)(1-\rho^2)^{-\frac{1}{2}} = y$ in the integrand of (4·4). $AV(\hat{\rho})$ can then be expressed as

$$AV(\hat{\rho}) = \frac{\rho^2(1-\rho^2)^2}{n}\left\{1 + \frac{1}{b_2 - b_1^2/b_0}\right\},$$

where $b_k = E'[X^k\phi(X)\phi(-X)]$, with E' denoting expectation with respect to the distribution of $\mathcal{N}[\omega(1-\rho^2)^{-\frac{1}{2}}, \rho^2(1-\rho^2)^{-1}]$. It can now be seen from considerations of symmetry

† Columns (from left to right) and rows (from top to bottom) correspond to ω, ρ, μ, σ.

that b_0 and b_2 have maxima at $\omega = 0$, while b_1 has its minimum at $\omega = 0$, which completes the proof.

5. THE UTILITY OF r^*

We now present a series of results concerning r^*, which will be followed by a general discussion of its value. Note that expression (3·2) for r^* is invariant under a linear transformation of the X_i, so in all results pertaining only to r^*, we set $\omega = 0$ and $\sigma^2 = 1$.

A little later we shall need $E(XZ)$. Since it is not difficult to obtain, we will give the expression for the general moment $\alpha_k = E(X^k Z)$.

THEOREM III.
$$\alpha_k = \sum_{j=0}^{k} \binom{k}{j} (1-\rho^2)^{\frac{1}{2}j} \rho^{k-j} a_j \int_{\omega}^{+\infty} y^{k-j} e^{-\frac{1}{2}y^2} dy,$$

where a_j is the j-th moment of the random variable $\mathcal{N}(0,1)$.

Proof. Using the definition of E_1, we obtain

$$E(X^k Z) = p E_1(X^k) = \int_{-\infty}^{+\infty} \int_{\omega}^{+\infty} x^k \psi(x,y) \, dy \, dx.$$

Make the transformation $t = (x - \rho y)\sqrt{(1-\rho^2)}$. The above then reduces to

$$\int_{-\infty}^{+\infty} \int_{\omega}^{+\infty} \{t\sqrt{(1-\rho^2)} + y\rho\}^k \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}t^2} \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}y^2} \, dy \, dt.$$

Using a binomial expansion and integrating with respect to t, we obtain Theorem III. The integrals which occur are incomplete gamma functions which may be evaluated by the usual recursion relation.

For completeness we include the relation between $\rho(X, Y)$ and $\rho(X, Z)$, known to Karl Pearson:

$$\rho(X, Z) = \rho(X, Y) \frac{\lambda(\omega)}{\sqrt{(pq)}}.$$

It follows from the original definition of biserial correlation, as given by Pearson, that r^* is consistent. This fact is also an immediate consequence of relation (3·2) between r^* and r: $r \to \rho(X, Z)$ in probability as $n \to \infty$. Thus

$$r^* = \frac{r}{\lambda(T)} \left\{ \frac{1}{n} \Sigma(Z_i - \bar{Z})^2 \right\}^{\frac{1}{2}} \to \frac{\rho(X, Z)\sqrt{(pq)}}{\lambda(\omega)} \quad \text{in probability,}$$

and hence by the above, $r^* \to \rho(X, Y)$ in probability.

With respect to the magnitude of r^*, it is well known that $|r^*|$ can be greater than 1. Something of the nature of this phenomenon can be understood by looking at r. In order to prove a result concerning the magnitude of r^*, we shall need a preliminary result (see Tate, 1953, Lemma 2):

THEOREM IV. $p(x) q(x) \geqslant \frac{1}{2}\pi\lambda^2(x)$, $(-\infty < x < +\infty)$, *with equality at 0, $\pm\infty$.*

Now we have

THEOREM V.
$$\frac{r^*}{r} \geqslant \sqrt{(\tfrac{1}{2}\pi)}.$$

Proof. Rewriting (3·2) as

$$r^* = r \frac{\sqrt{(\bar{Z} - \bar{Z}^2)}}{\lambda(T)},$$

we have, in view of the definition of T,
$$r^* = r \frac{\sqrt{\{p(T) q(T)\}}}{\lambda(T)}.$$

Theorem IV applies for any T, so Theorem V is proved. As a consequence of Theorem V, we see that

$$r^* \genfrac{}{}{0pt}{}{> +1}{< -1} \quad \text{according as} \quad r \genfrac{}{}{0pt}{}{> + \sqrt{\frac{2}{\pi}}}{< - \sqrt{\frac{2}{\pi}}}.$$

Asymptotic normality of r^*, which will be needed later in this section, is a consequence of a theorem of Cramér.

THEOREM VI. $r^* \sim \mathcal{N}\{\rho, AV(r^*)\}$.

Proof. In expression (3·2) the term $\lambda(T)$ is seen to be an infinitely differentiable function of $\bar{Z}$. Thus, r^* is a totally differentiable function of the sample means $\bar{X}, \bar{Z}, \bar{X}^2, \overline{XZ}$. Applying Cramér's theorem (see Cramér, 1946, p. 366), we have asymptotic normality with the asymptotic variance (3·4) calculated by Soper.

We shall now present two results which are more important than those just preceding. They concern the asymptotic, or large-sample, efficiency of r^*, with respect to the class of estimators of ρ based on the sample (X_i, Z_i).

THEOREM VII. *r^* is an asymptotically most efficient estimator of ρ when $\rho = 0$.*

Proof. In view of Theorem VI on asymptotic normality, we have a right to inquire about the asymptotic efficiency of r^*, which will be denoted by

$$AEff(r^*) = \frac{AV(\hat{\rho})}{AV(r^*)}.$$

It may be seen from Theorem I that
$$AV(\hat{\rho} \mid \omega, 0) = \frac{p(\omega) q(\omega)}{n\{\lambda(\omega)\}^2}. \tag{5·1}$$

Now, from (3·4) we observe that (5·1) coincides with $AV(r^* \mid \omega, 0)$. The conclusion follows from the definition of an asymptotically most efficient estimator.

THEOREM VIII. *r^* is an asymptotically least efficient estimator of ρ when $|\rho| \to 1$.*

Proof. An application of Theorem IV shows that

$$\phi\left(\frac{\omega - \rho x}{\sqrt{(1-\rho^2)}}\right) \phi\left(-\frac{\omega - \rho x}{\sqrt{(1-\rho^2)}}\right) \leqslant \frac{2}{\pi}.$$

Hence, recalling the definition of $g(x, \omega, \rho)$ in Theorem I, we see that all integrals of the form $\int_{-\infty}^{+\infty} x^k g(x, \omega, \rho)\, dx$ exist. Schwarz's inequality shows that $AV(\hat{\rho} \mid \omega, \rho)$ is such that the term in braces is non-vanishing. Thus, $AV(\hat{\rho} \mid \omega, \rho) \to 0$ as $|\rho| \to 1$. From the fact $AV(r^* \mid \omega, \rho) \to \frac{2}{\pi}$ as $|\rho| \to 1$, we conclude that $AEff(r^* \mid \omega, \rho) \to 0$.

The special case $\omega = 0$ has interesting features which will appear in Theorems X and XI. First we shall need another preliminary result (see Tate, 1953, Lemma 1):

THEOREM IX. $\{1 - 2p(x)\} \lambda(x) - xp(x) q(x) \geqslant 0, \quad (x \geqslant 0)$.

Theorem X. *The asymptotic variance of r^* has its minimum for each ρ at $\omega = 0$.*

Proof. In view of symmetry, it will be sufficient to show the result for $\omega \geqslant 0$. Let

$$A(\omega) = \frac{\omega^2 p(\omega) q(\omega)}{\{\lambda(\omega)\}^2} - \frac{\omega\{1-2p(\omega)\}}{\lambda(\omega)}, \quad B(\omega) = \frac{p(\omega) q(\omega)}{\{\lambda(\omega)\}^2},$$

$$g(\omega) = \{1-2p(\omega)\}\lambda(\omega) - \omega p(\omega) q(\omega), \quad h(\omega) = p(\omega) q(\omega) - \pi\{\lambda(\omega)\}^2/2.$$

From this point until the end of the proof, we shall omit ω whenever it appears as an argument of any function. We have (Tate, 1953)

$$g' = 2\lambda^2 - pq, \quad g'' = -4\omega\lambda^2 + (2q-1)\lambda, \quad g(0) = g(\infty) = 0, \quad g'(0) > 0,$$

$$h' = \lambda(1-2q) + \pi\omega\lambda^2, \quad h'' = \lambda^2(\pi-2-2\pi\omega^2) - \omega(1-2q)\lambda,$$

$$h(0) = h(\infty) = h'(0) = 0, \quad h''(0) > 0.$$

Accordingly, we have

$$A = -\omega g\lambda^{-2}, \quad B = h\lambda^{-2} + \tfrac{1}{2}\pi, \quad \text{with} \quad A \leqslant 0, \quad B \geqslant \tfrac{1}{2}\pi,$$

both equalities holding at $\omega = 0$. The relation $AV(r^* \mid \omega, \rho) \geqslant AV(r^* \mid 0, \rho)$ for all ρ may be written $\rho^2 A + B \geqslant \tfrac{1}{2}\pi$ for all ρ. Since $A \leqslant 0$ this last expression is implied by $A + B \geqslant \tfrac{1}{2}\pi$, which in turn is equivalent to $h \geqslant \omega g$. Thus, we must show $k = h - \omega g \geqslant 0$:

$$
\left.
\begin{aligned}
k' &= 2\omega q(1-q) - 2(2q-1)\lambda + \omega(\pi-2)\lambda^2, \\
k'' &= 2q(1-q) - \lambda^2\{6-\pi+\omega^2(2\pi-4)\}, \\
k(0) &= k(\infty) = k'(0) = 0, \quad k''(0) = 1 - 3/\pi > 0.
\end{aligned}
\right\}
\tag{5·2}
$$

We shall show that there exists no y such that $k'(y) = 0$, $k(y) < 0$. Suppose such a y does exist. Then

$$
\left.
\begin{aligned}
2(2q-1)\lambda &= 2yq(1-q) + (\pi-2)y\lambda^2, \\
q(1-q)(1+y^2) - \tfrac{1}{2}\pi\lambda^2 - y(2q-1)\lambda &< 0.
\end{aligned}
\right\}
\tag{5·3}
$$

Substituting the right member of the first expression into the second, we have

$$2q(1-q) < \lambda^2\{\pi + (\pi-2)y^2\}.$$

Thus, $k''(y) < \lambda^2\{2\pi - 6 - (\pi-2)y^2\}$. A negative maximum must, from (5·2), be followed by a negative minimum. Hence, from the above relation in $k''(y)$, there exist no extrema which exceed $\{(2\pi-6)/(\pi-2)\}^{\frac{1}{2}}$. Assuming there is a negative extremum of k, then there must be a negative minimum in $(0, 1)$. Let y be this minimum point. Then $k''(y) > 0$, or from (5·2), $2q(1-q) - \lambda^2\{6-\pi-(2\pi-4)y^2\} > 0$. Substituting the value of $2q(1-q)$ obtained from the first equation in (5·3), we reach $(2q-1) - y\lambda\{2 + (\pi-2)y^2\} > 0$. The left member vanishes at $y = 0$ and has a negative derivative for $0 \leqslant y^2 \leqslant 1$. Therefore, there is no negative minimum in $(0, 1)$, and from the previous argument $k \geqslant 0$, which completes the proof.

Since for any fixed ρ, r^* is a better estimator when $\omega = 0$, it will be useful to have for this case something simpler in the way of an asymptotic distribution of r^* than that contained in Theorem VI. We are therefore led to

Theorem XI. *When $\omega = 0$, we have to a close approximation*

$$\tanh^{-1}\frac{2r^*}{\sqrt{5}} \sim \mathcal{N}\left(\tanh^{-1}\frac{2\rho}{\sqrt{5}}, \frac{5}{4n}\right).$$

Proof.

$$AV(r^* \mid 0, \rho) = \frac{1}{n}(\rho^4 - \tfrac{5}{2}\rho^2 + \tfrac{1}{4}\pi) = \frac{1}{n}(\tfrac{1}{4} - \rho^2)^2 - \left(\frac{25 - 8\pi}{16n}\right).$$

Dropping the last term and solving the equation

$$\sqrt{g'(x)} = \frac{1}{(\tfrac{1}{4} - x^2)},$$

we have $g(x) = (2\sqrt{5})\tanh^{-1}(2x\sqrt{5})$. It is known that

$$\sqrt{n}\{g(r^*) - g(\rho)\} \sim \mathcal{N}(0, 1),$$

so the theorem is proved.

Discussion of results concerning r^*

In looking over Theorems V, VI, VII, VIII, X and XI, several facts stand out. First, even though r^* is consistent and asymptotically normal, it is still inadequate for estimating ρ because of its possible magnitude and its lack of large sample efficiency for large values of $|\rho|$. In the case of testing the hypothesis $H : \rho = \rho_0$ the first defect is not of so much consequence. Even in a problem of estimation, one can always operate under the rule: when $|r^*| < 1$, estimate ρ by r^*; when $r^* > 1$, estimate $\rho = 1$; and when $r^* < -1$, estimate $\rho = -1$. The gross defect is lack of efficiency. In practically all applications it is of more interest to detect large values of ρ than small values. In just such cases r^* is a 'worst' estimator. On the other hand, again speaking in large-sample terms, when $\rho = 0$, r^* is a 'best' estimator.[†] Hence, if we base a test of $H : \rho = \rho_0$ on r^*, good results should be achieved when $|\rho_0|$ is small. It is then recommended that r^* be used for one and only one purpose, to test $H : \rho = \rho_0$ when $|\rho_0|$ is small. If, in addition, the assumption $\omega = 0$ is tenable, then the variance stabilizing transformation of Theorem XI may be used, calculations being performed with Table V B of Fisher (1946, p. 210). In such a case certain advantages (see Fisher 1946, pp. 197–204) will accrue. $\{nAV(r^*)\}^{\frac{1}{2}}$ is given in Table 2. Note that from Theorem X, $\omega = 0$ is desirable on the grounds of precision.

6. SOLUTION OF THE LIKELIHOOD EQUATIONS

Using the notation of (4·1), we have the likelihood equations

$$\delta L = \Sigma \left\{ z_i \frac{\delta\xi\left(\frac{x_i - \mu}{\sigma}, \omega\right)}{\xi\left(\frac{x_i - \mu}{\sigma}, \omega\right)} + (1 - z_i) \frac{\delta\eta\left(\frac{x_i - \mu}{\sigma}, \omega\right)}{\eta\left(\frac{x_i - \mu}{\sigma}, \omega\right)} \right\} = 0. \tag{6·1}$$

As was mentioned previously (cf. footnote, p. 208), the solution of the four-parameter problem reduces to the problem of determining $\hat{\omega}$ and $\hat{\rho}$. It turns out that the likelihood equations for μ and ω may be combined to yield $\hat{\mu} = \bar{x}$. Similarly, a combination of likelihood equations for σ^2 and ρ will give us $\hat{\sigma}^2 = s^2$. Details will be omitted.

We now replace μ and σ by $\bar{x}$ and s in the expression $(x_i - \mu)/\sigma$, which occurs in (6·1), and denote the result by x_i'. Also, let L' denote the new likelihood function. The solution of $\delta L' = 0$ for $\hat{\omega}$ and $\hat{\rho}$ follows.

[†] Another point in favour of r^* is indicated by the parallelism of Theorems II and X.

One may easily verify that

$$\frac{\partial \eta(x'_i, \omega)}{\partial \omega} = \psi(x'_i, \omega). \qquad\qquad \frac{\partial \xi(x'_i, \omega)}{\partial \omega} = -\psi(x'_i, \omega).$$

$$\frac{\partial \eta(x'_i, \omega)}{\partial \rho} = \frac{-(x'_i - \rho\omega)\,\psi(x'_i, \omega)}{(1-\rho^2)}, \quad \frac{\partial \xi(x'_i, \omega)}{\partial \rho} = \frac{(x'_i - \rho\omega)\,\psi(x'_i, \omega)}{(1-\rho^2)}. \qquad (6\cdot2)$$

By the use of (6·2) the equations $\delta L' = 0$ can be written as

$$\Sigma(x'_i - \rho\omega)(2z_i - 1)\phi\left\{(2z_i - 1)\left(\frac{\omega - \rho x'_i}{\sqrt{(1-\rho^2)}}\right)\right\} = 0,$$

$$\Sigma(2z_i - 1)\phi\left\{(2z_i - 1)\left(\frac{\omega - \rho x'_i}{\sqrt{(1-\rho^2)}}\right)\right\} = 0. \qquad (6\cdot3)$$

Now introduce the notation

$$\delta_i = 2z_i - 1, \quad \gamma_i = (\omega - \rho x'_i)(1-\rho^2)^{-\frac{1}{2}}, \quad \phi_i = \phi(\delta_i\gamma_i),$$

$$A_i = \phi_i(\phi_i - \delta_i\gamma_i). \qquad (6\cdot4)$$

Rewriting (6·3) again, in the new notation, we have

$$\Sigma\delta_i\phi_i = 0, \quad \Sigma\delta_i x'_i\phi_i = 0. \qquad (6\cdot5)$$

Easy differentiation gives $\phi'(x) = \phi(x)\{\phi(x) - x\}$. Newton's method in two variables gives the following equations in $\Delta\omega$ and $\Delta\rho$, where $\Delta\omega = \omega - \omega_1$, $\Delta\rho = \rho - \rho_1$, ω_1 and ρ_1 being initial guesses:

$$\left(\Sigma\frac{A_i}{\sqrt{(1-\rho^2)}}\right)\Delta\omega + \left(\frac{\rho_1\omega_1\Sigma A_i - \Sigma A_i x'_i}{(1-\rho_1^2)^{\frac{3}{2}}}\right)\Delta\rho = -\Sigma\delta_i\phi_i.$$

$$\left(\Sigma\frac{A_i x'_i}{\sqrt{(1-\rho^2)}}\right)\Delta\omega + \left(\frac{\rho_1\omega_1\Sigma A_i x'_i - \Sigma A_i x'^2_i}{(1-\rho_1^2)^{\frac{3}{2}}}\right)\Delta\rho = -\Sigma\delta_i x'_i\phi_i. \qquad (6\cdot6)$$

Let Δ be the determinant of the coefficients. The method of solution will then be the following:

(i) Compute ω^*, r^* from the sample (x_i, z_i) $(i = 1, 2, \ldots, n)$, where r^* is the sample biserial correlation coefficient and ω^* is the solution of the equation $p(\omega) = \bar{z}$. Now, let $\omega_1 = \omega^*$ and

$$\rho_1 = \begin{cases} r^* & \text{when} & |r^*| < 1, \\ +0\cdot90 & \text{when} & r^* \geqslant 1, \\ -0\cdot90 & \text{when} & r^* \leqslant -1. \end{cases}$$

(ii) Compute δ_i, γ_i, ϕ_i, $\delta_i\phi_i$, $\delta_i x'_i\phi_i$, A_i, $A_i x'_i$, $A_i x'^2_i$ for $i = 1, 2, \ldots, n$, where δ_i, γ_i, ϕ_i, A_i are defined in (6·4), and the numerical values of the $\phi_i = \phi(\delta_i\gamma_i)$ are obtained from the tables printed on pp. 217–221 below. Note that these tables must be entered with $X = \delta_i\gamma_i$, while $\phi_i = Z/Q$ if $X > 0$ and Z/P if $X < 0$.

(iii) Evaluate the three determinants

$$\Delta = \begin{vmatrix} \Sigma A_i & \rho_1\omega_1\Sigma A_i - \Sigma A_i x'_i \\ \Sigma A_i x'_i & \rho_1\omega_1\Sigma A_i x'_i - \Sigma A_i x'^2_i \end{vmatrix} \frac{1}{(1-\rho_1^2)^2},$$

$$\Delta\omega = \begin{vmatrix} -\Sigma\delta_i\phi_i & \rho_1\omega_1\Sigma A_i - \Sigma A_i x'_i \\ -\Sigma\delta_i x'_i\phi_i & \rho_1\omega_1\Sigma A_i x'_i - \Sigma A_i x'^2_i \end{vmatrix} \frac{1}{\Delta(1-\rho_1^2)^{\frac{3}{2}}},$$

$$\Delta\rho = \begin{vmatrix} \Sigma A_i & -\Sigma\delta_i\phi_i \\ \Sigma A_i x'_i & -\Sigma\delta_i x'_i\phi_i \end{vmatrix} \frac{1}{\Delta(1-\rho_1^2)^{\frac{1}{2}}}.$$

(iv) Obtain (ω, ρ) from $(\Delta\omega, \Delta\rho)$ and (ω_1, ρ_1) and repeat the process using $\omega = \omega_2, \rho = \rho_2$ in place of ω_1 and ρ_1.

The rule given in (i) is somewhat arbitrary, but is believed to be a good rule of thumb. The longest stage in the scheme outlined above is the determination of ϕ_i, $i = 1, 2, \ldots, n$, from the tables.

We shall now present an illustration of the method. In order to have a good vantage point for observing the way the calculations run, we select a random sample from a bivariate normal population with zero means, unit variances and correlation $1/\sqrt{2}$. A table of random numbers from such a population is not available directly†, but can be constructed from a table of random numbers from $\mathcal{N}(0, 1)$ as follows: Let $U = \mathcal{N}(0, 1)$, $V = \mathcal{N}(0, 1)$, $\omega = \frac{1}{2}$, with U and V independent. Now let $X = U$ and $Y = (U + V)/\sqrt{2}$. After introducing the dichotomy

$$Z = 1 \text{ if } Y \geqslant \tfrac{1}{2}, \quad Z = 0 \text{ if } Y < \tfrac{1}{2}, \tag{6.7}$$

we have the desired set-up. It is important to note that for this particular example, the x_i' of the computing scheme become merely x_i, since it is known that $\mu = 0$ and $\sigma^2 = 1$. For computations see Table 1 below.

A second iteration resulted in $\omega_3 = 0\cdot251$, $\rho_3 = 0\cdot489$. Since $\hat{\rho}$ remained unchanged in the third decimal, the results were not included. Recall that the true value of $\hat{\rho}$ is $0\cdot707$. On the basis of our sample of 20, $\hat{\rho} = 0\cdot489$ is the best we can do. However, by using the iterative scheme instead of r^* we removed 27 % of the error.

Although r^* is used merely to start the iteration, the fact that

$$|r^* - E(r^*)|/\sqrt{\{AV(r^*)\}} = 1\cdot64$$

serves to indicate that we could have been more fortunate in our selection of a sample.

REFERENCES

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

DU BOIS, P. H. (1942). A note on the computation of biserial r in item validation. *Psychometrika*, **7**, 143.

DUNLAP, J. W. (1936). A nomograph for computing biserial r. *Psychometrika*, **1**, 59.

FISHER, R. A. (1946). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

MARITZ, J. S. (1953). Estimation of the correlation coefficient of a bivariate normal population where one of the variables is dichotomized. *Psychometrika*, **18**, 97.

PEARSON, K. (1909). On a new method for determining the correlation between a measured character A and a character B. *Biometrika*, **7**, 96.

ROYER, E. B. (1941). Punched cards methods for determining biserial coefficients. *Psychometrika*, **6**, 55.

SOPER, H. E. (1913). On the probable error for the biserial expression for the correlation coefficient. *Biometrika*, **10**, 384.

TATE, R. F. (1953). On a double inequality of the normal distribution. *Ann. Math. Statist.* **24**, 133.

TOCHER, K. D. (1949). A note on the analysis of grouped probit data. *Biometrika*, **36**, 9.

† [A table of *Correlated Random Normal Deviates* is now at press and will be issued shortly from the Department of Statistics, University College, London as *Tracts for Computers* No. XXVI. ED.]

Table 1. *Computation of example*

i	x_i	z_i	γ_i	δ_i	ϕ_i	$\delta_i x_i \phi_i$	$\phi_i - \delta_i \gamma_i$	A_i	$A_i x_i$	$A_i x_1^2$
1	0·24	0	0·030	—	0·779	0·187	0·809	0·630	0·151	0·036
2	0·63	1	−0·146		0·707	0·445	0·853	0·603	0·380	0·239
3	−0·59	0	0·404	—	0·560	−0·330	0·964	0·540	−0·319	0·188
4	1·08	0	−0·348	—	1·032	1·115	0·684	0·706	0·762	0·822
5	0·06	0	0·111	—	0·729	0·044	0·840	0·451	0·027	0·002
6	−0·01	0	0·143	—	0·709	−0·007	0·852	0·604	−0·006	0·000
7	1·59	1	−0·578		0·470	0·747	1·048	0·493	0·784	1·247
8	−0·41	0	0·323	—	0·604	−0·248	0·927	0·560	−0·230	0·094
9	−0·33	1	0·287		0·989	−0·326	0·702	0·694	−0·229	0·076
10	0·99	1	−0·308		0·613	0·607	0·921	0·565	0·559	0·553
11	0·30	0	0·003	—	0·796	0·239	0·799	0·636	0·191	0·057
12	−2·07	0	1·070	—	0·262	−0·542	1·332	0·349	−0·722	1·495
13	−0·21	0	0·233	—	0·655	−0·138	0·888	0·582	−0·122	0·026
14	−0·47	1	0·350		1·033	0·486	0·683	0·706	−0·332	0·156
15	1·28	1	−0·438		0·541	0·692	0·979	0·530	0·678	0·868
16	0·82	1	−0·231		0·656	0·538	0·887	0·582	0·477	0·391
17	1·06	1	−0·339		0·595	0·631	0·934	0·556	0·589	0·624
18	1·09	0	−0·353	—	1·035	1·128	0·682	0·706	0·770	0·829
19	0·57	0	−0·119	—	0·875	0·499	0·756	0·662	0·377	0·215
20	−0·53	0	0·377	—	1·052	−0·558	0·675	0·710	−0·376	0·199

$$\Sigma x_i = 5·09 \qquad \Sigma x_i z_i = 5·04 \qquad r^* = +0·410\,(\rho_1) \qquad \gamma_i = 0·138 - 0·450 x_i$$
$$\Sigma x_1^2 = 15·626 \qquad \Sigma z_i = 9 \qquad \omega^* = +0·126\,(\omega_1)$$
$$\Sigma \delta_i \phi_i = -1·380 \qquad \Sigma A_i = 11·865 \quad \Sigma A_i x_1^2 = 8·127 \qquad \Delta = -122·54$$
$$\Sigma \delta_i x_i \phi_i = 0·343 \qquad \Sigma A_i x_i = 3·409 \qquad \Delta\omega = 0·108 \qquad \Delta\rho = 0·079$$
$$\omega_2 = 0·234 \qquad \rho_2 = 0·489$$

Table 2. *The asymptotic standard deviation of r^* (biserial r) as a function of p and ρ (equation (3·4))*

All values must be divided by $\sqrt{n}$, as the quantity tabled is $\{nA\,V(r^*)\}^{\frac{1}{2}}$

					p or $1-p$					
	0·05	0·10	0·15	0·20	0·25	0·30	0·35	0·40	0·45	0·50
0·00	4·466	2·922	2·345	2·041	1·857	1·737	1·658	1·608	1·580	1·571
0·10	2·104	1·699	1·521	1·419	1·353	1·308	1·278	1·258	1·247	1·243
0·20	2·077	1·668	1·491	1·389	1·323	1·279	1·248	1·228	1·217	1·213
0·30	2·033	1·616	1·440	1·339	1·273	1·229	1·198	1·179	1·167	1·163
0·40	1·971	1·543	1·370	1·269	1·203	1·159	1·128	1·109	1·097	1·093
0·50	1·893	1·449	1·279	1·179	1·114	1·069	1·038	1·019	1·008	1·004
0·60	1·799	1·333	1·167	1·069	1·004	0·960	0·930	0·910	0·898	0·894
0·70	1·691	1·194	1·034	0·939	0·875	0·831	0·801	0·781	0·769	0·766
0·80	1·569	1·031	0·881	0·789	0·727	0·683	0·653	0·632	0·620	0·616
0·90	1·438	0·842	0·705	0·619	0·559	0·517	0·486	0·465	0·453	0·449
1·00	1·302	0·616	0·503	0·429	0·374	0·335	0·304	0·283	0·270	0·266

$\pm\rho$ (row labels at left)

THE NORMAL PROBABILITY FUNCTION: TABLES OF CERTAIN AREA-ORDINATE RATIOS AND OF THEIR RECIPROCALS

EDITORIAL

Notation. Writing X as a standardized normal deviate, the following notation will be used:[*]

$$Z = Z(X) = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}X^2},$$

$$P = P(X) = \int_{-\infty}^{X}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}u^2}\,du, \quad Q = 1 - P = \int_{X}^{\infty}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}u^2}\,du.$$

The table then gives for $X = 0{\cdot}00\,(0{\cdot}01)\,3{\cdot}00$ the values of the four ratios P/Z, Q/Z, Z/P and Z/Q. For the last three ratios, the results are extended for argument $X = 3{\cdot}00\,(0{\cdot}01)\,4{\cdot}00\,(0{\cdot}05)\,5{\cdot}00$. The ratios P/Z, Q/Z and Z/Q are tabled to five decimal places and those for Z/P to five significant figures.

Derivation. The ratio P/Z was freshly computed in the Department of Statistics, University College, London, using the *Tables of Normal Probability Functions*, National Bureau of Standards (1953). The ratio Q/Z has been taken directly from Table III of Karl Pearson's *Tables for Statisticians and Biometricians*, Part II (1931). The ratios Z/P and Z/Q were calculated by Prof. Z. W. Birnbaum using the National Bureau of Standards Tables, and were originally included in two tables forming part of the preceding paper by Dr R. F. Tate (1955). With Dr Tate's and Prof. Birnbaum's approval it was, however, decided to give separately the more comprehensive table which follows. This will form a useful companion to Table II, *Tables for Statisticians and Biometricians*, Part II, which gives the same four ratios for an argument of P instead of X. It is intended to reproduce both tables in the second volume of *Biometrika Tables for Statisticians*.

Having regard to this wider use, Dr Tate's notation has been modified, so that his $\phi(x)$ (see p. 206 above) becomes Z/P for $x = -X < 0$ and Z/Q for $x = X > 0$.

Interpolation. For a large part of the table accuracy to the full number of decimal places given may be obtained by linear interpolation. Where this fails, the Bessel formula

$$y_\theta = (1-\theta)\,y_0 + \theta y_1 - \tfrac{1}{4}\theta(1-\theta)\,(\delta^2 y_0 + \delta^2 y_1)$$

is adequate, except for P/Z with $X > 2{\cdot}50$, when full use of second differences in Everett's formula becomes necessary.

The normal probability function; extension of main table

X	Q/Z	Z/P	Z/Q	X	Q/Z	Z/P	Z/Q
4·00	0·23665	$0{\cdot}0^3 13383$	4·22561	4·50	0·21257	$0{\cdot}0^4 15984$	4·70432
·05	·23401	$\cdot 0^3 10944$	4·27330	·55	·21042	$\cdot 0^4 12747$	4·75240
·10	·23143	$\cdot 0^4 89264$	4·32103	·60	·20831	$\cdot 0^4 10141$	4·80051
·15	·22890	$\cdot 0^4 72627$	4·36880	·65	·20624	$\cdot 0^5 80472$	4·84865
·20	·22642	$\cdot 0^4 58944$	4·41662	·70	·20421	$\cdot 0^5 63698$	4·89682
4·25	0·22399	$0{\cdot}0^4 47719$	4·46447	4·75	0·20222	$0{\cdot}0^5 50295$	4·94503
·30	·22161	$\cdot 0^4 38536$	4·51237	·80	·20027	$\cdot 0^5 39613$	4·99326
·35	·21928	$\cdot 0^4 31042$	4·56030	·85	·19835	$\cdot 0^5 31122$	5·04153
·40	·21700	$\cdot 0^4 24943$	4·60827	·90	·19647	$\cdot 0^5 24390$	5·08983
·45	·21476	$\cdot 0^4 19992$	4·65628	·95	·19462	$\cdot 0^5 19066$	5·13815
4·50	0·21257	$0{\cdot}0^4 15984$	4·70432	5·00	0·19281	$0{\cdot}0^5 14867$	5·18650

[*] This conforms to the notation used by Pearson and Hartley in *Biometrika Tables for Statisticians*, 1 (1954).

The normal probability function. Table of area/ordinate ratios and their reciprocals

X	P/Z	Q/Z	Z/P	Z/Q	X	P/Z	Q/Z	Z/P	Z/Q
0·00	1·25331	1·25331	0·79788	0·79788	**0·50**	1·96402	0·87636	0·50916	1·14108
·01	1·26338	1·24338	·79153	·80426	·51	1·98399	·87078	·50404	1·14840
·02	1·27357	1·23356	·78520	·81066	·52	2·00426	·86525	·49894	1·15574
·03	1·28389	1·22387	·77888	·81708	·53	2·02483	·85977	·49387	1·16310
·04	1·29434	1·21430	·77260	·82352	·54	2·04572	·85436	·48882	1·17047
0·05	1·30492	1·20484	0·76633	0·82999	**0·55**	2·06693	0·84900	0·48381	1·17786
·06	1·31564	1·19550	·76008	·83647	·56	2·08846	·84370	·47882	1·18526
·07	1·32650	1·18627	·75386	·84298	·57	2·11032	·83845	·47386	1·19268
·08	1·33750	1·17716	·74766	·84950	·58	2·13252	·83326	·46893	1·20011
·09	1·34864	1·16816	·74149	·85605	·59	2·15506	·82812	·46402	1·20756
0·10	1·35993	1·15926	0·73533	0·86262	**0·60**	2·17795	0·82303	0·45915	1·21503
·11	1·37136	1·15048	·72920	·86921	·61	2·20120	·81799	·45430	1·22251
·12	1·38295	1·14179	·72309	·87582	·62	2·22481	·81301	·44948	1·23000
·13	1·39468	1·13321	·71701	·88244	·63	2·24879	·80807	·44468	1·23751
·14	1·40658	1·12474	·71095	·88910	·64	2·27315	·80319	·43992	1·24504
0·15	1·41862	1·11636	0·70491	0·89577	**0·65**	2·29789	0·79835	0·43518	1·25258
·16	1·43083	1·10809	·69889	·90246	·66	2·32302	·79357	·43047	1·26013
·17	1·44320	1·09991	·69290	·90917	·67	2·34855	·78883	·42579	1·26770
·18	1·45574	1·09183	·68694	·91590	·68	2·37449	·78414	·42114	1·27529
·19	1·46844	1·08384	·68099	·92265	·69	2·40085	·77949	·41652	1·28289
0·20	1·48132	1·07594	0·67507	0·92942	**0·70**	2·42763	0·77489	0·41192	1·29050
·21	1·49437	1·06814	·66918	·93620	·71	2·45484	·77034	·40736	1·29813
·22	1·50760	1·06043	·66331	·94301	·72	2·48249	·76583	·40282	1·30577
·23	1·52101	1·05281	·65746	·94984	·73	2·51059	·76137	·39831	1·31342
·24	1·53460	1·04527	·65164	·95669	·74	2·53915	·75695	·39383	1·32109
0·25	1·54837	1·03782	0·64584	0·96355	**0·75**	2·56817	0·75257	0·38938	1·32878
·26	1·56234	1·03046	·64007	·97044	·76	2·59767	·74824	·38496	1·33648
·27	1·57650	1·02318	·63432	·97734	·77	2·62766	·74394	·38057	1·34419
·28	1·59085	1·01599	·62859	·98426	·78	2·65814	·73969	·37620	1·35191
·29	1·60541	1·00887	·62290	·99121	·79	2·68913	·73548	·37187	1·35965
0·30	1·62017	1·00184	0·61722	0·99817	**0·80**	2·72063	0·73131	0·36756	1·36740
·31	1·63513	0·99488	·61157	1·00514	·81	2·75266	·72718	·36328	1·37517
·32	1·65030	·98801	·60595	1·01214	·82	2·78523	·72309	·35904	1·38295
·33	1·66569	·98121	·60035	1·01916	·83	2·81835	·71904	·35482	1·39074
·34	1·68130	·97448	·59478	1·02619	·84	2·85202	·71503	·35063	1·39854
0·35	1·69713	0·96783	0·58923	1·03324	**0·85**	2·88626	0·71106	0·34647	1·40636
·36	1·71318	·96126	·58371	1·04031	·86	2·92109	·70712	·34234	1·41419
·37	1·72946	·95475	·57821	1·04739	·87	2·95651	·70322	·33824	1·42204
·38	1·74598	·94832	·57274	1·05450	·88	2·99254	·69935	·33416	1·42989
·39	1·76273	·94196	·56730	1·06162	·89	3·02918	·69553	·33012	1·43776
0·40	1·77973	0·93567	0·56188	1·06876	**0·90**	3·06646	0·69173	0·32611	1·44564
·41	1·79697	·92944	·55649	1·07591	·91	3·10438	·68798	·32212	1·45354
·42	1·81447	·92329	·55113	1·08308	·92	3·14296	·68425	·31817	1·46144
·43	1·83222	·91720	·54579	1·09028	·93	3·18222	·68057	·31425	1·46936
·44	1·85023	·91118	·54047	1·09748	·94	3·22216	·67691	·31035	1·47729
0·45	1·86850	0·90522	0·53519	1·10471	**0·95**	3·26280	0·67329	0·30648	1·48524
·46	1·88704	·89932	·52993	1·11195	·96	3·30416	·66971	·30265	1·49319
·47	1·90586	·89349	·52470	1·11921	·97	3·34624	·66615	·29884	1·50116
·48	1·92496	·88772	·51949	1·12648	·98	3·38908	·66263	·29506	1·50914
·49	1·94434	·88201	·51431	1·13377	·99	3·43268	65914	·29132	1·51713
0·50	1·96402	0·87636	0·50916	1·14108	**1·00**	3·47705	0·65568	0·28760	1·52514

$$Z(X) = e^{-\frac{1}{2}X^2}/\sqrt{(2\pi)}, \quad P(X) = 1 - Q(X) = \int_{-\infty}^{X} Z(u)\, du.$$

X	P/Z	Q/Z	Z/P	Z/Q	X	P/Z	Q/Z	Z/P	Z/Q
1·00	3·47705	0·65568	0·28760	1·52514	1·50	7·20514	0·51582	0·13879	1·93868
·01	3·52222	·65225	·28391	1·53315	·51	7·32448	·51358	·13653	1·94719
·02	3·56821	·64885	·28025	1·54118	·52	7·44636	·51133	·13429	1·95570
·03	3·61502	·64549	·27662	1·54922	·53	7·57087	·50911	·13208	1·96423
·04	3·66268	·64215	·27302	1·55726	·54	7·69805	·50690	·12990	1·97276
1·05	3·71121	0·63885	0·26945	1·56532	1·55	7·82799	0·50472	0·12775	1·98130
·06	3·76062	·63557	·26591	1·57340	·56	7·96074	·50255	·12562	1·98985
·07	3·81094	·63232	·26240	1·58148	·57	8·09639	·50040	·12351	1·99841
·08	3·86218	·62910	·25892	1·58958	·58	8·23500	·49826	·12143	2·00698
·09	3·91437	·62591	·25547	1·59768	·59	8·37664	·49614	·11938	2·01555
1·10	3·96752	0·62274	0·25205	1·60580	1·60	8·52140	0·49404	0·11735	2·02413
·11	4·02166	·61961	·24865	1·61392	·61	8·66935	·49195	·11535	2·03272
·12	4·07681	·61650	·24529	1·62206	·62	8·82058	·48988	·11337	2·04131
·13	4·13299	·61342	·24196	1·63021	·63	8·97517	·48782	·11142	2·04992
·14	4·19023	·61036	·23865	1·63837	·64	9·13320	·48578	·10949	2·05853
1·15	4·24854	0·60733	0·23538	1·64654	1·65	9·29477	0·48376	0·10759	2·06715
·16	4·30795	·60433	·23213	1·65472	·66	9·45996	·48175	·10571	2·07578
·17	4·36849	·60135	·22891	1·66292	·67	9·62887	·47975	·10385	2·08441
·18	4·43018	·59840	·22572	1·67112	·68	9·80159	·47777	·10202	2·09305
·19	4·49305	·59548	·22257	1·67933	·69	9·97824	·47580	·10022	2·10170
1·20	4·55713	0·59257	0·21944	1·68755	1·70	10·15889	0·47385	0·098436	2·11036
·21	4·62243	·58970	·21634	1·69578	·71	10·34367	·47192	·096677	2·11902
·22	4·68900	·58684	·21326	1·70403	·72	10·53268	·46999	·094943	2·12769
·23	4·75685	·58402	·21022	1·71228	·73	10·72604	·46808	·093231	2·13637
·24	4·82603	·58121	·20721	1·72054	·74	10·92384	·46619	·091543	2·14506
1·25	4·89655	0·57843	0·20422	1·72882	1·75	11·12622	0·46431	0·089878	2·15375
·26	4·96845	·57567	·20127	1·73710	·76	11·33330	·46244	·088236	2·16245
·27	5·04176	·57294	·19834	1·74539	·77	11·54520	·46058	·086616	2·17115
·28	5·11652	·57022	·19544	1·75369	·78	11·76205	·45874	·085019	2·17987
·29	5·19276	·56754	·19258	1·76201	·79	11·98397	·45692	·083445	2·18859
1·30	5·27051	0·56487	0·18974	1·77033	1·80	12·21112	0·45510	0·081893	2·19731
·31	5·34980	·56222	·18692	1·77866	·81	12·44362	·45330	·080362	2·20605
·32	5·43069	·55960	·18414	1·78700	·82	12·68163	·45151	·078854	2·21479
·33	5·51319	·55699	·18138	1·79535	·83	12·92528	·44973	·077368	2·22353
·34	5·59735	·55441	·17866	1·80371	·84	13·17474	·44797	·075903	2·23229
1·35	5·68321	0·55185	0·17596	1·81208	1·85	13·43017	0·44622	0·074459	2·24105
·36	5·77081	·54931	·17329	1·82046	·86	13·69171	·44448	·073037	2·24982
·37	5·86019	·54679	·17064	1·82884	·87	13·95955	·44275	·071636	2·25859
·38	5·95139	·54430	·16803	1·83724	·88	14·23386	·44104	·070255	2·26737
·39	6·04446	·54182	·16544	1·84564	·89	14·51481	·43934	·068895	2·27615
1·40	6·13944	0·53936	0·16288	1·85406	1·90	14·80258	0·43765	0·067556	2·28495
·41	6·23638	·53692	·16035	1·86248	·91	15·09737	·43597	·066237	2·29375
·42	6·33533	·53450	·15784	1·87091	·92	15·39937	·43430	·064938	2·30255
·43	6·43632	·53210	·15537	1·87935	·93	15·70877	·43265	·063659	2·31136
·44	6·53942	·52972	·15292	1·88780	·94	16·02580	·43100	·062399	2·32018
1·45	6·64467	0·52735	0·15050	1·89626	1·95	16·35065	0·42937	0·061160	2·32900
·46	6·75213	·52501	·14810	1·90473	·96	16·68354	·42775	·059939	2·33784
·47	6·86186	·52268	·14573	1·91320	·97	17·02472	·42614	·058738	2·34667
·48	6·97389	·52038	·14339	1·92169	·98	17·37440	·42454	·057556	2·35551
·49	7·08830	·51809	·14108	1·93018	·99	17·73282	·42295	·056393	2·36436
1·50	7·20514	0·51582	0·13879	1·93868	2·00	18·10025	0·42137	0·055248	2·37322

The normal probability function. Table of area/ordinate ratios and their reciprocals (cont.)

X	P/Z	Q/Z	Z/P	Z/Q	X	P/Z	Q/Z	Z/P	Z/Q
2·00	18·10025	0·42137	0·055248	2·37322	2·50	56·69633	0·35427	0·017638	2·82274
·01	18·47692	·41980	·054122	2·38208	·51	58·14464	·35313	·017198	2·83186
·02	18·86311	·41825	·053014	2·39094	·52	59·63565	·35199	·016768	2·84097
·03	19·25908	·41670	·051924	2·39981	·53	61·17075	·35087	·016348	2·85010
·04	19·66512	·41516	·050851	2·40869	·54	62·75138	·34975	·015936	2·85922
2·05	20·08152	0·41364	0·049797	2·41758	2·55	64·37902	0·34863	0·015533	2·86835
·06	20·50857	·41212	·048760	2·42646	·56	66·05523	·34753	·015139	2·87748
·07	20·94657	·41062	·047741	2·43536	·57	67·78159	·34643	·014753	2·88662
·08	21·39586	·40912	·046738	2·44426	·58	69·55976	·34533	·014376	2·89576
·09	21·85675	·40764	·045752	2·45317	·59	71·39146	·34425	·014007	2·90491
2·10	22·32959	0·40616	0·044784	2·46208	2·60	73·27844	0·34316	0·013647	2·91406
·11	22·81471	·40470	·043831	2·47100	·61	75·22256	·34209	·013294	2·92321
·12	23·31249	·40324	·042896	2·47992	·62	77·22570	·34102	·012949	2·93237
·13	23·82329	·40179	·041976	2·48885	·63	79·28985	·33996	·012612	2·94153
·14	24·34749	·40036	·041072	2·49778	·64	81·41704	·33890	·012282	2·95070
2·15	24·88550	0·39893	0·040184	2·50672	2·65	83·60939	0·33785	0·011960	2·95987
·16	25·43771	·39751	·039312	2·51566	·66	85·86908	·33681	·011646	2·96904
·17	26·00455	·39610	·038455	2·52461	·67	88·19839	·33577	·011338	2·97822
·18	26·58645	·39470	·037613	2·53357	·68	90·59968	·33474	·011038	2·98740
·19	27·18387	·39331	·036787	2·54253	·69	93·07536	·33371	·010744	2·99658
2·20	27·79726	0·39193	0·035975	2·55150	2·70	95·62799	0·33269	0·010457	3·00577
·21	28·42711	·39055	·035178	2·56047	·71	98·26016	·33168	·010177	3·01497
·22	29·07391	·38919	·034395	2·56944	·72	100·97461	·33067	·0²99035	3·02416
·23	29·73816	·38783	·033627	2·57842	·73	103·77414	·32967	·0²96363	3·03336
·24	30·42041	·38649	·032873	2·58741	·74	106·66167	·32867	·0²93754	3·04257
2·25	31·12118	0·38515	0·032132	2·59640	2·75	109·64022	0·32768	0·0²91207	3·05177
·26	31·84105	·38382	·031406	2·60540	·76	112·71295	·32669	·0²88721	3·06098
·27	32·58060	·38250	·030693	2·61440	·77	115·88308	·32571	·0²86294	3·07020
·28	33·34041	·38118	·029994	2·62341	·78	119·15401	·32474	·0²83925	3·07942
·29	34·12113	·37988	·029307	2·63242	·79	122·52923	·32377	·0²81613	3·08864
2·30	34·92338	0·37858	0·028634	2·64144	2·80	126·01238	0·32280	0·0²79357	3·09787
·31	35·74783	·37729	·027974	2·65046	·81	129·60721	·32184	·0²77156	3·10710
·32	36·59516	·37601	·027326	2·65948	·82	133·31763	·32089	·0²75009	3·11633
·33	37·46608	·37474	·026691	2·66851	·83	137·14770	·31994	·0²72914	3·12556
·34	38·36133	·37348	·026068	2·67755	·84	141·10162	·31900	·0²70871	3·13480
2·35	39·28165	0·37222	0·025457	2·68659	2·85	145·18375	0·31806	0·0²68878	3·14405
·36	40·22783	·37097	·024858	2·69563	·86	149·39863	·31713	·0²66935	3·15329
·37	41·20068	·36973	·024271	2·70468	·87	153·75095	·31620	·0²65040	3·16254
·38	42·20103	·36850	·023696	2·71374	·88	158·24559	·31528	·0²63193	3·17180
·39	43·22974	·36727	·023132	2·72280	·89	162·88761	·31436	·0²61392	3·18106
2·40	44·28771	0·36605	0·022580	2·73186	2·90	167·68228	0·31345	0·0²59637	3·19032
·41	45·37586	·36484	·022038	2·74093	·91	172·63504	·31254	·0²57926	3·19958
·42	46·49515	·36364	·021508	2·75000	·92	177·75156	·31164	·0²56258	3·20885
·43	47·64656	·36244	·020988	2·75908	·93	183·03773	·31074	·0²54634	3·21812
·44	48·83112	·36125	·020479	2·76816	·94	188·49965	·30985	·0²53050	3·22739
2·45	50·04988	0·36007	0·019980	2·77725	2·95	194·14366	0·30896	0·0²51508	3·23667
·46	51·30393	·35889	·019492	2·78634	·96	199·97636	·30808	·0²50006	3·24595
·47	52·59442	·35773	·019013	2·79543	·97	206·00458	·30720	·0²48543	3·25523
·48	53·92250	·35657	·018545	2·80453	·98	212·23545	·30632	·0²47118	3·26452
·49	55·28938	·35541	·018087	2·81364	·99	218·67633	·30545	·0²45730	3·27381
2·50	56·69633	0·35427	0·017638	2·82274	3·00	225·33490	0·30459	0·0²44378	3·28310

$$Z(X) = e^{-\frac{1}{2}X^2}/\sqrt{(2\pi)}, \quad P(X) = 1 - Q(X) = \int_{-\infty}^{X} Z(u)\, du.$$

X	Q/Z	Z/P	Z/Q	X	Q/Z	Z/P	Z/Q
3·00	0·30459	$0.0^2 44378$	3·28310	3·50	0·26657	$0.0^3 87289$	3·75139
·01	·30373	$.0^2 43063$	3·29240	·51	·26590	$.0^3 84281$	3·76082
·02	·30287	$.0^2 41782$	3·30169	·52	·26523	$.0^3 81370$	3·77026
·03	·30202	$.0^2 40535$	3·31100	·53	·26457	$.0^3 78551$	3·77969
·04	·30118	$.0^2 39322$	3·32030	·54	·26391	$.0^3 75822$	3·78913
3·05	0·30034	$0.0^2 38141$	3·32961	3·55	0·26326	$0.0^3 73180$	3·79857
·06	·29950	$.0^2 36992$	3·33892	·56	·26260	$.0^3 70624$	3·80802
·07	·29866	$.0^2 35874$	3·34824	·57	·26195	$.0^3 68150$	3·81746
·08	·29784	$.0^2 34787$	3·35755	·58	·26131	$.0^3 65756$	3·82691
·09	·29701	$.0^2 33729$	3·36687	·59	·26066	$.0^3 63440$	3·83636
3·10	0·29619	$0.0^2 32700$	3·37620	3·60	0·26002	$0.0^3 61200$	3·84581
·11	·29538	$.0^2 31699$	3·38552	·61	·25939	$.0^3 59033$	3·85527
·12	·29456	$.0^2 30726$	3·39485	·62	·25875	$.0^3 56936$	3·86472
·13	·29376	$.0^2 29780$	3·40418	·63	·25812	$.0^3 54909$	3·87418
·14	·29295	$.0^2 28860$	3·41352	·64	·25749	$.0^3 52949$	3·88364
3·15	0·29215	$0.0^2 27965$	3·42286	3·65	0·25686	$0.0^3 51053$	3·89311
·16	·29136	$.0^2 27096$	3·43220	·66	·25624	$.0^3 49221$	3·90257
·17	·29057	$.0^2 26251$	3·44154	·67	·25562	$.0^3 47449$	3·91204
·18	·28978	$.0^2 25430$	3·45089	·68	·25500	$.0^3 45737$	3·92151
·19	·28900	$.0^2 24632$	3·46024	·69	·25439	$.0^3 44082$	3·93098
3·20	0·28822	$0.0^2 23857$	3·46959	3·70	0·25378	$0.0^3 42482$	3·94046
·21	·28744	$.0^2 23104$	3·47895	·71	·25317	$.0^3 40937$	3·94993
·22	·28667	$.0^2 22373$	3·48830	·72	·25256	$.0^3 39444$	3·95941
·23	·28590	$.0^2 21662$	3·49767	·73	·25196	$.0^3 38002$	3·96889
·24	·28514	$.0^2 20972$	3·50703	·74	·25136	$.0^3 36608$	3·97838
3·25	0·28438	$0.0^2 20302$	3·51640	3·75	0·25076	$0.0^3 35263$	3·98786
·26	·28363	$.0^2 19652$	3·52576	·76	·25017	$.0^3 33963$	3·99735
·27	·28287	$.0^2 19020$	3·53514	·77	·24957	$.0^3 32708$	4·00683
·28	·28213	$.0^2 18407$	3·54451	·78	·24898	$.0^3 31496$	4·01632
·29	·28138	$.0^2 17812$	3·55389	·79	·24840	$.0^3 30326$	4·02582
3·30	0·28064	$0.0^2 17234$	3·56327	3·80	0·24781	$0.0^3 29197$	4·03531
·31	·27990	$.0^2 16674$	3·57265	·81	·24723	$.0^3 28107$	4·04481
·32	·27917	$.0^2 16130$	3·58203	·82	·24665	$.0^3 27054$	4·05431
·33	·27844	$.0^2 15602$	3·59142	·83	·24607	$.0^3 26039$	4·06381
·34	·27772	$.0^2 15090$	3·60081	·84	·24550	$.0^3 25059$	4·07331
3·35	0·27699	$0.0^2 14593$	3·61020	3·85	0·24493	$0.0^3 24114$	4·08281
·36	·27627	$.0^2 14112$	3·61960	·86	·24436	$.0^3 23202$	4·09232
·37	·27556	$.0^2 13644$	3·62900	·87	·24379	$.0^3 22322$	4·10183
·38	·27485	$.0^2 13191$	3·63840	·88	·24323	$.0^3 21474$	4·11134
·39	·27414	$.0^2 12752$	3·64780	·89	·24267	$.0^3 20656$	4·12085
3·40	0·27343	$0.0^2 12326$	3·65720	3·90	0·24211	$0.0^3 19866$	4·13036
·41	·27273	$.0^2 11914$	3·66661	·91	·24155	$.0^3 19106$	4·13988
·42	·27203	$.0^2 11513$	3·67602	·92	·24100	$.0^3 18372$	4·14940
·43	·27134	$.0^2 11126$	3·68544	·93	·24045	$.0^3 17665$	4·15892
·44	·27065	$.0^2 10750$	3·69485	·94	·23990	$.0^3 16983$	4·16844
3·45	0·26996	$0.0^2 10386$	3·70427	3·95	0·23935	$0.0^3 16326$	4·17796
·46	·26927	$.0^2 10033$	3·71369	·96	·23881	$.0^3 15693$	4·18749
·47	·26859	$.0^3 96911$	3·72311	·97	·23826	$.0^3 15083$	4·19702
·48	·26791	$.0^3 93601$	3·73254	·98	·23772	$.0^3 14495$	4·20654
·49	·26724	$.0^3 90394$	3·74196	·99	·23719	$.0^3 13929$	4·21608
3·50	0·26657	$0.0^3 87289$	3·75139	4·00	·0·23665	$0.0^3 13383$	4·22561

REFERENCES

NATIONAL BUREAU OF STANDARDS (1953). *Tables of Normal Probability Functions*, Applied Mathematics Series, no. 23. Washington: U.S. Department of Commerce.

PEARSON, E. S. & HARTLEY, H. O. (1954). *Biometrika Tables for Statisticians*, **1**. Cambridge University Press for the Biometrika Trustees.

PEARSON, K. (1931). *Tables for Statisticians and Biometricians*, **2**. Cambridge University Press for the Biometrika Trustees.

TATE, R. F. (1955). The theory of correlation between two continuous variables when one is dichotomized. *Biometrika*, **42**, 205.

TABLES OF SYMMETRIC FUNCTIONS. PART V

By F. N. DAVID and M. G. KENDALL

1. We write M for monomial symmetric functions, S for the one-part functions, U for the unitary functions and H for the homogeneous product sums. Previously we have given tables MS-SM (1949), UM-MU, HU-UH (1951) and MH-HM (1953). We now complete the set, up to and including weight (w) 12, by the US-SU tables. The HS-SH are the same as the present tables as far as coefficients are concerned.

2. The US-SU tables have been partially calculated when in the course of constructing the previous sets. There are many ways in which they can be built up. We have used here both the D operator technique and the elementary relations

$$n!\,a_n = \Sigma(-1)^{n+\pi_1+\pi_2+\cdots+\pi_k}\frac{n!}{1^{\pi_1}2^{\pi_2}\ldots k^{\pi_k}\pi_1!\,\pi_2!\ldots\pi_k!}\,s_1^{\pi_1}s_2^{\pi_2}\ldots s_k^{\pi_k},$$

$$S_n = \Sigma(-1)^{n+\pi_1+\cdots+\pi_k}\frac{(\pi_1+\pi_2+\cdots+\pi_k-1)!\,n}{\pi_1!\,\pi_2!\ldots\pi_k!}\,a_1^{\pi_1}a_2^{\pi_2}\ldots a_k^{\pi_k},$$

the sum in each case being taken over all possible partitions.

3. To express the S-functions in terms of the U-functions we read horizontally up to and including the diagonal figure in bold type. Thus for $w = 5$ we have, for example,

$$(3)(2) = a_1^5 - 5a_2a_1^3 + 6a_2^2a_1 + 3a_3a_1^2 - 6a_3a_2.$$

To express the U-functions in terms of the S-functions we read vertically downwards up to and including the diagonal in bold type and divide the coefficients by $w!$. Thus

$$a_3a_2 \cdot 5! = 10(1)^5 - 40(2)(1)^3 + 30(2)^2(1) + 20(3)(1)^2 - 20(3)(2).$$

4. We have called attention previously to the use of symmetric functions in distribution problems. These present tables will be useful for that purpose. They will also add considerably to the flexibility of the symmetric function system, for with this concluding set it will now be possible to express any symmetric function in terms of any other.

REFERENCES

David, F. N. & Kendall, M. G. (1949). *Biometrika*, **36**, 431.

David, F. N. & Kendall, M. G. (1951). *Biometrika*, **38**, 435.

David, F. N. & Kendall, M. G. (1953). *Biometrika*, **40**, 427.

Table 5.2

$w=2$	a_1^2	a_2
$(1)^2$	$\dfrac{2}{1}$	1
(2)	1	$\dfrac{-1}{-2}$

Table 5.3

$w=3$	a_1^3	a_2a_1	a_3
$(1)^3$	$\dfrac{6}{1}$	3	1
$(2)(1)$	1	$\dfrac{-3}{-2}$	-3
(3)	1	-3	$\dfrac{2}{3}$

Table 5.4

$w=4$	a_1^4	$a_2a_1^2$	a_2^2	a_3a_1	a_4
$(1)^4$	$\dfrac{24}{1}$	12	6	4	1
$(2)(1)^2$	1	$\dfrac{-12}{-2}$	-12	-12	-6
$(2)^2$	1	-4	$\dfrac{6}{4}$	.	3
$(3)(1)$	1	-3	.	$\dfrac{8}{3}$	8
(4)	1	-4	2	4	$\dfrac{-6}{-4}$

Table 5.5

$w=5$	a_1^5	$a_2a_1^3$	$a_2^2a_1$	$a_3a_1^2$	a_3a_2	a_4a_1	a_5
$(1)^5$	$\dfrac{120}{1}$	60	30	20	10	5	1
$(2)(1)^3$	1	$\dfrac{-60}{-2}$	-60	-60	-40	-30	-10
$(2)^2(1)$	1	-4	$\dfrac{30}{4}$	.	30	15	15
$(3)(1)^2$	1	-3	.	$\dfrac{40}{3}$	20	40	20
$(3)(2)$	1	-5	6	3	$\dfrac{-20}{-6}$	.	-20
$(4)(1)$	1	-4	2	4	.	$\dfrac{-30}{-4}$	-30
(5)	1	-5	5	5	-5	-5	$\dfrac{24}{5}$

Table 5.6

$w=6$	a_1^6	$a_2a_1^4$	$a_2^2a_1^2$	a_2^3	$a_3a_1^3$	$a_3a_2a_1$	a_3^2	$a_4a_1^2$	a_4a_2	a_5a_1	a_6
$(1)^6$	$\dfrac{720}{1}$	360	180	90	120	60	20	30	15	6	1
$(2)(1)^4$	1	$\dfrac{-360}{-2}$	-360	-270	-360	-240	-120	-180	-105	-60	-15
$(2)^2(1)^2$	1	-4	$\dfrac{180}{4}$	270	.	180	180	90	135	90	45
$(2)^3$	1	-6	12	$\dfrac{-90}{-8}$	.	.	.	.	-45	.	15
$(3)(1)^3$	1	-3	.	.	$\dfrac{240}{3}$	120	80	240	120	120	40
$(3)(2)(1)$	1	-5	6	.	3	$\dfrac{-120}{-6}$	-240	.	-120	-120	-120
$(3)^2$	1	-6	9	.	6	-18	$\dfrac{80}{9}$	.	.	.	40
$(4)(1)^2$	1	-4	2	.	4	.	.	$\dfrac{-180}{-4}$	-90	-180	-90
$(4)(2)$	1	-6	10	-4	4	-8	.	-4	$\dfrac{90}{8}$	.	90
$(5)(1)$	1	-5	5	.	5	-5	.	-5	.	$\dfrac{144}{5}$	144
(6)	1	-6	9	-2	6	-12	3	-6	6	6	$\dfrac{-120}{-6}$

Table 5.7

$w=7$	a_1^7	$a_2a_1^5$	$a_2^2a_1^3$	$a_2^3a_1$	$a_3a_1^4$	$a_3a_2a_1^2$	$a_3a_2^2$	$a_3^2a_1$	$a_4a_1^3$	$a_4a_2a_1$	a_4a_3	$a_5a_1^2$	a_5a_2	a_6a_1	a_7
$(1)^7$	$\dfrac{5040}{1}$	2520	1260	630	840	420	210	140	210	105	35	42	21	7	1
$(2)(1)^5$	1	$\dfrac{-2520}{-2}$	-2520	-1890	-2520	-1680	-1050	-840	-1260	-735	-315	-420	-231	-105	-21
$(2)^2(1)^3$	1	-4	$\dfrac{1260}{4}$	1890	.	1260	1470	1260	630	945	735	630	525	315	105
$(2)^3(1)$	1	-6	12	$\dfrac{-630}{-8}$	.	.	-630	.	.	-315	-315	.	-315	-105	-105
$(3)(1)^4$	1	-3	.	.	$\dfrac{1680}{3}$	840	420	560	1680	840	350	840	420	280	70
$(3)(2)(1)^2$	1	-5	6	.	3	$\dfrac{-840}{-6}$	-840	-1680	.	-840	-1260	-840	-840	-840	-420
$(3)(2)^2$	1	-7	16	-12	3	-12	$\dfrac{420}{12}$	.	.	.	210	.	420	.	210
$(3)^2(1)$	1	-6	9	.	6	-18	.	$\dfrac{560}{9}$	.	.	560	.	.	280	280
$(4)(1)^3$	1	-4	2	.	4	.	.	.	$\dfrac{-1260}{-4}$	-630	-210	-1260	-630	-630	-210
$(4)(2)(1)$	1	-6	10	-4	4	-8	.	.	-4	$\dfrac{630}{8}$	630	.	630	630	630
$(4)(3)$	1	-7	14	-6	7	-24	6	12	-4	12	$\dfrac{-420}{-12}$	.	.	.	-420
$(5)(1)^2$	1	-5	5	.	5	-5	.	.	-5	.	.	$\dfrac{1008}{5}$	504	1008	504
$(5)(2)$	1	-7	13	-10	5	-15	10	.	-5	10	.	5	$\dfrac{-504}{-10}$	.	-504
$(6)(1)$	1	-6	9	-2	6	-12	.	3	-6	6	.	6	.	$\dfrac{-840}{-6}$	-840
(7)	1	-7	14	-7	7	-21	7	7	-7	14	-7	7	-7	-7	$\dfrac{720}{7}$

Table 5.8

$w=8$ (i)	a_1^8	$a_2a_1^6$	$a_2^2a_1^4$	$a_2^3a_1^2$	a_2^4	$a_3a_1^5$	$a_3a_2a_1^3$	$a_3a_2^2a_1$	$a_3^2a_1^2$	$a_3^2a_2$	$a_4a_1^4$
$(1)^8$	$\frac{40320}{1}$	20160	10080	5040	2520	6720	3360	1680	1120	560	1680
$(2)(1)^6$	1	$\frac{-20160}{-2}$	20160	-15120	10080	20160	13440	-8400	6720	-3920	-10080
$(2)^2(1)^4$	1	-4	$\frac{10080}{4}$	15120	15120	.	10080	11760	10080	8400	5040
$(2)^3(1)^2$	1	-6	12	$\frac{-5040}{-8}$	-10080	.	.	-5040	.	-5040	.
$(2)^4$	1	-8	24	-32	$\frac{2520}{16}$	.	.	.	.	.	.
$(3)(1)^5$	1	-3	.	.	.	$\frac{13440}{3}$	6720	3360	4480	2240	13440
$(3)(2)(1)^3$	1	-5	6	.	.	3	$\frac{-6720}{-6}$	-6720	-13440	-8960	.
$(3)(2)^2(1)$	1	-7	16	-12	.	3	-12	$\frac{3360}{12}$	.	6720	.
$(3)^2(1)^2$	1	-6	9	.	.	6	-18	.	$\frac{4480}{9}$	2240	.
$(3)^2(2)$	1	-8	21	-18	.	6	-30	36	9	$\frac{-2240}{-18}$	.
$(4)(1)^4$	1	-4	2	-4	.	4	-8	.	.	.	$\frac{-10080}{-4}$
$(4)(2)(1)^2$	1	-6	10	-4	.	4	-16	16	.	.	-4
$(4)(2)^2$	1	-8	22	-24	8	4	-24	6	12	.	-4
$(4)(3)(1)$	1	-7	14	-6	.	7	-32	16	16	.	-8
$(4)^2$	1	-8	20	-16	4	8	.	.	.	.	.
$(5)(1)^3$	1	-5	5	.	.	5	-5	.	.	.	-5
$(5)(2)(1)$	1	-7	15	-10	.	5	-15	10	.	.	-5
$(5)(3)$	1	-8	20	-15	.	8	-35	30	15	-15	-6
$(6)(1)^2$	1	-6	9	-2	.	6	-12	3	3	-6	-6
$(6)(2)$	1	-8	21	-20	4	6	-24	24	7	7	-7
$(7)(1)$	1	-7	14	-7	.	7	-21	7	12	-8	-8
(8)	1	-8	20	-16	2	8	-32	24			

$w=8$ (ii)	$a_4a_3a_1^2$	$a_4a_3^2$	$a_4a_3a_1$	a_4^2	$a_5a_1^3$	$a_5a_2a_1$	a_5a_3	$a_6a_1^2$	a_6a_2	a_7a_1	a_8
	840	420	280	70	336	168	56	56	28	8	1
$(1)^8$	-5880	-3360	-2520	-840	-3360	-1848	-728	-840	-448	-168	-28
$(2)(1)^6$	7560	6720	5880	2940	5040	4200	2520	2520	1680	840	210
$(2)^2(1)^4$	-2520	-5040	-2520	-2520	.	-2520	-2520	-840	-1680	-840	-420
$(2)^3(1)^2$	.	1200	.	.	630	.	.	.	420	.	105
$(2)^4$	6720	3360	2800	1120	6720	3360	1232	2240	1120	560	112
$(3)(1)^5$	-6720	-6720	-10080	-6720	-6720	-6720	-5600	-6720	-4480	-3360	-1120
$(3)(2)(1)^3$	.	3360	1680	3360	.	3360	5040	.	3360	1680	1680
$(3)(2)^2(1)$	.	.	.	4480	.	.	3360	2240	1120	2240	1120
$(3)^2(1)^2$	.	.	4480	4480	.	.	.	-2240	.	.	-1120
$(3)^2(2)$	.	.	.	-840	-10080	-5040	1680	-5040	-2520	-1680	-420
$(4)(1)^4$	-5040	-2520	-1680	-840	-10080	-5040	1680	-5040	5040	5040	2520
$(4)(2)(1)^2$	$\frac{5040}{8}$	5040	5040	5040	.	5040	5040	5040	-2520	.	-1260
$(4)(2)^2$	16	$\frac{-2520}{-16}$	.	-2520	.	.	.	.	-3360	-3360	-3360
$(4)(3)(1)$	12	.	$\frac{-3360}{-12}$	-6720	.	.	-3360	.	.	.	1260
$(4)^2$	32	-16	-32	$\frac{2520}{16}$	8064	4032	1344	8064	4032	4032	1344
$(5)(1)^3$	.	.	.	.	$\frac{8064}{5}$	-4032	-4032	.	-4032	-4032	-4032
$(5)(2)(1)$	10	.	.	.	5	$\frac{-4032}{-10}$	-4032	.	.	.	2688
$(5)(3)$	15	.	-15	.	5	-15	$\frac{2688}{15}$	-6720	-3360	-6720	-3360
$(6)(1)^2$	6	.	.	.	6	.	.	$\frac{-6720}{-6}$	3360	.	3360
$(6)(2)$	18	-12	.	.	6	-12	.	-6	$\frac{3360}{12}$	5760	5760
$(7)(1)$	14	.	-7	.	7	-7	.	-7	7	$\frac{5760}{7}$	-5040
(8)	24	-8	-16	4	8	-16	8	-8	8	8	$\frac{-5040}{-8}$

Table 5.9

$w = 9$ (i)

$w=9$ (i)	a_1^9	$a_2a_1^7$	$a_2^2a_1^5$	$a_2^3a_1^3$	$a_2^4a_1$	$a_3a_1^6$	$a_3a_2a_1^4$	$a_3a_2^2a_1^2$	$a_3a_2^3$	$a_3^2a_1^3$	$a_3^2a_2a_1$	a_3^3	$a_4a_1^5$	$a_4a_2a_1^3$	$a_4a_2^2a_1$
$(1)^9$	**362880** / 1	181440	90720	45360	22680	60480	30240	15120	7560	10080	5040	1680	15120	7560	378
$(2)(1)^7$	1	**−181440** / −2	−181440	136080	−90720	−181440	−120960	−75600	45360	60480	−35280	−15120	90720	32720	30240
$(2)^2(1)^6$	1	−4	**90720** / 4	136080	136080	.	90720	105840	90720	90720	75600	45360	45360	68040	60480
$(2)^3(1)^3$	1	−6	12	**−45360** / −8	−90720	.	.	−45360	−75600	.	−45360	−45360	.	−22680	45360
$(2)^4(1)$	1	−8	24	−32	**22680** / 16	.	.	.	.	22680	.	.	.	.	11340
$(3)(1)^6$	1	−3	.	.	.	**120960** / 3	60480	30240	15120	40320	20160	10080	120960	60480	3024
$(3)(2)(1)^4$	1	−5	6	.	.	3	**−60480** / −6	−60480	−45360	−120960	−80640	−60480	.	−60480	6048
$(3)(2)^2(1)^2$	1	−7	16	−12	.	3	−12	**30240** / 12	45360	.	60480	90720	.	.	3024
$(3)(2)^3$	1	−9	30	−44	24	3	−18	36	**−15120** / −24	.	.	.	.	.	.
$(3)^2(1)^3$	1	−6	9	.	.	6	−18	.	.	**40320** / 9	20160	20160	.	.	.
$(3)^2(2)(1)$	1	−8	21	−18	.	6	−30	36	.	9	**−20160** / −18	−60480	.	.	.
$(3)^3$	1	−9	27	−27	.	9	−54	81	.	27	−81	**13440** / 27	.	.	.
$(4)(1)^5$	1	−4	2	.	.	4	.	.	.	.	.	.	**−90720** / −4	−45360	22680
$(4)(2)(1)^3$	1	−6	10	−4	.	4	−8	.	.	.	.	.	−4	**45360** / 8	45360
$(4)(2)^2(1)$	1	−8	22	−24	8	4	−16	16	.	.	.	.	−4	16	**−22680** / −16
$(4)(3)(1)^2$	1	−7	14	−6	.	7	−24	6	.	12	.	.	−4	12	.
$(4)(3)(2)$	1	−9	28	−34	12	7	−38	54	−12	12	−24	.	−4	20	−24
$(4)^2(1)$	1	−8	20	−10	4	8	−32	16	.	16	.	.	−8	32	−16
$(5)(1)^4$	1	−5	5	.	.	5	−5	.	.	.	.	.	−5	10	.
$(5)(2)(1)^2$	1	−7	15	−10	.	5	−15	10	.	.	.	.	−5	20	−20
$(5)(2)^2$	1	−9	29	−40	20	5	−25	40	−20	.	.	.	−5	15	.
$(5)(3)(1)$	1	−8	20	−15	.	8	−35	30	.	15	−15	.	−5	40	.
$(5)(4)$	1	−9	27	−30	10	9	−45	50	−10	20	−20	.	−9	.	.
$(6)(1)^3$	1	−6	9	−2	.	6	−12	.	.	3	.	.	−6	6	.
$(6)(2)(1)$	1	−8	21	−20	4	6	−24	24	.	3	−6	.	−6	18	12
$(6)(3)$	1	−9	27	−29	6	9	−48	63	−6	21	−45	9	−6	24	−18
$(7)(1)^2$	1	−7	14	−7	.	7	−21	7	.	7	.	.	−7	28	−28
$(7)(2)$	1	−9	28	−35	14	7	−35	49	−14	7	−14	.	−8	24	−8
$(8)(1)$	1	−8	20	−16	2	8	−32	24	.	12	−8	.	−9	36	−27
(9)	1	−9	27	−30	9	9	−45	54	−9	18	−27	3	.	.	.

$w = 9$ (ii)

$w=9$ (ii)	$a_4a_3a_1^2$	$a_4a_3a_2$	$a_4^2a_1$	$a_5a_1^4$	$a_5a_2a_1^2$	$a_5a_2^2$	$a_5a_3a_1$	a_5a_4	$a_6a_1^3$	$a_6a_2a_1$	a_6a_3	$a_7a_1^2$	a_7a_2	a_8a_1	a_9
$(1)^9$	2520	1260	630	3024	1512	756	504	126	504	252	84	72	36	9	1
$(2)(1)^7$	−22680	−12600	−7560	−30240	−16632	−9072	6552	−2016	−7560	−4032	−1512	−1512	−792	252	36
$(2)^2(1)^6$	52920	37800	26460	45360	37800	27216	22680	9828	22680	15120	7560	7560	4536	1890	378
$(2)^3(1)^3$	−22680	−37800	−22680	.	−22680	−30240	−22680	−15120	−7560	−15120	−12600	−7560	−7560	−3780	−1260
$(2)^4(1)$	.	11340	5670	.	.	11340	.	5670	.	3780	3780	.	3780	945	945
$(3)(1)^6$	25200	12600	10080	60480	30240	15120	11088	3528	20160	10080	3528	5040	2520	1008	168
$(3)(2)(1)^4$	90720	−57960	−60480	−60480	−60480	45360	50400	−27720	60480	−40320	−22680	−30240	17640	10080	2520
$(3)(2)^2(1)^2$	15120	52920	30240	.	30240	45360	45360	37800	.	30240	37800	15120	22680	15120	7560
$(3)(2)^3$	.	−7560	.	.	.	−15120	.	−7560	.	.	−2520	.	−7560	.	−2520
$(3)^2(1)^3$	40320	20160	40320	.	.	.	20160	20160	20160	10080	10080	20160	10080	10080	3360
$(3)^2(2)(1)$	.	−20160	.	.	.	.	.	−20160	−20160	.	−10080	−30240	.	−10080	−10080
$(3)^3$	.	.	.	.	.	.	.	.	.	.	6720	.	.	.	2240
$(4)(1)^5$	−15120	−7560	−7560	−90720	−45360	−22680	−15120	4536	45360	−22680	7560	15120	−7560	3780	−756
$(4)(2)(1)^3$	45360	30240	45360	.	45360	45360	45360	30240	45360	45360	30240	45360	30240	22680	7560
$(4)(2)^2(1)$	.	−22680	−22680	.	.	−22680	.	−22680	.	−22680	−22680	.	−22680	−11340	−11340
$(4)(3)(1)^2$	−30240 / −12	−15120	−60480	.	.	.	−30240	−45360	.	.	−15120	−30240	−15120	−30240	−15120
$(4)(3)(2)$	−12	**15120** / 24	.	.	.	.	.	15120	.	15120	.	15120	.	15120	15120
$(4)^2(1)$	−32	.	**22680** / 16	.	.	.	.	22680	.	.	.	.	.	11340	11340
$(5)(1)^4$	.	.	.	**72576** / 5	36288	18144	12096	3024	72576	36288	12096	36288	18144	12096	3024
$(5)(2)(1)^2$	.	.	.	5	**−36288** / −10	−36288	−36288	−18144	.	−36288	−36288	−36288	−36288	−36288	−18144
$(5)(2)^2$	.	.	.	5	−20	**18144** / 20	.	9072	.	.	.	.	.	18144	9072
$(5)(3)(1)$	−15	.	.	5	−15	.	**24192** / 15	24192	.	24192	.	.	.	24192	24192
$(5)(4)$	−40	20	20	5	−30	10	20	**−18144** / −20	.	.	.	.	.	.	−18144
$(6)(1)^3$	.	.	.	6	.	.	.	.	**−60480** / −6	−30240	−10080	−60480	−30240	−30240	−10080
$(6)(2)(1)$	.	.	.	6	−12	.	.	.	−6	**30240** / 12	30240	.	30240	30240	30240
$(6)(3)$	−18	18	.	6	−18	.	18	.	−6	18	**−20160** / −18	.	.	.	−20160
$(7)(1)^2$	−7	.	.	7	−7	.	.	.	−7	.	.	**51840** / 7	25920	51840	25920
$(7)(2)$	−7	14	.	7	−21	14	.	.	−7	14	.	7	**−25920** / −14	.	−25920
$(8)(1)$	−16	.	4	8	−16	.	8	.	−8	8	.	.	.	**−45360** / −8	45360
(9)	−27	18	9	9	−27	9	18	−9	−9	18	−9	9	−9	−9	**40320** / 9

Table 5.10

$m=10$ (i)	a_1^{10}	$a_2 a_1^8$	$a_2^2 a_1^6$	$a_2^3 a_1^4$	$a_2^4 a_1^2$	a_2^5	$a_3 a_1^7$	$a_3 a_2 a_1^5$	$a_3 a_2^2 a_1^3$	$a_3 a_2^3 a_1$
$(1)^{10}$	3628800	1814400	907200	453600	226800	113400	604800	302400	151200	75600
	1	−1814400	−1814400	1360800	907200	567000		−1814400	−1209600	−756000	−453600
$(2)(1)^8$	1	−2	907200	1360800	1360800	1134000			907200	1058400	907200
$(2)^2(1)^6$	1	−4	4	−453600	−907200	−1134000				−453600	−756000
$(2)^3(1)^4$	1	−6	12	−8	226800	567000					226800
$(2)^4(1)^2$	1	−8	24	−32	16	−113400					
$(2)^5$	1	−10	40	−80	80	−32	1209600	604800	302400	151200
$(3)(1)^7$	1	−3	·	·	·	·	3	−604800	−604800	−453600
$(3)(2)(1)^5$	1	−5	6	·	·	·	3	−6	302400	453600
$(3)(2)^2(1)^3$	1	−7	16	−12	·	·	3	−12	12	−151200
$(3)(2)^3(1)$	1	−9	30	−44	24	·	3	−18	36	−24
$(3)^2(1)^4$	1	−6	9	−18	·	·	6	−18	36	
$(3)^2(2)(1)^2$	1	−8	21	−18	·	·	6	−30	96	−72
$(3)^2(2)^2$	1	−10	37	−60	36	·	6	−42	81	
$(3)^3(1)$	1	−9	27	−27	·	·	9	−54		
$(4)(1)^6$	1	−4	2	·	·	·	4	−6	·	·
$(4)(2)(1)^4$	1	−6	10	−4	8	·	4	−16	16	−32
$(4)(2)^2(1)^2$	1	−8	22	−24	56	−16	4	−24	48	−12
$(4)(2)^3$	1	−10	38	−68			7	−24	6	−36
$(4)(3)(1)^3$	1	−7	14	−6	12		7	−38	54	
$(4)(3)(2)(1)$	1	−9	28	−34	18		10	−66	120	−32
$(4)(3)^2$	1	−10	35	−48	4	−8	8	−32	16	
$(4)^2(1)^2$	1	−8	20	−16	36		8	−48	80	−32
$(4)^2(2)$	1	−10	36	−56	36	−8				
$(5)(1)^5$	1	−5	5	−10	·	·	5	−5	10	−20
$(5)(2)(1)^3$	1	−7	15	−40	20	·	5	−15	40	
$(5)(2)^2(1)$	1	−9	29	−15	·	·	8	−25	30	−60
$(5)(3)(1)^2$	1	−8	20	−55	30	·	8	−51	100	−10
$(5)(3)(2)$	1	−10	36	−30	10	·	9	−45	50	−50
$(5)(4)(1)$	1	−9	27	−50	25	·	10	−60	100	
$(5)^2$	1	−10	35					−12		
$(6)(1)^4$	1	−6	9	−2	·	·	6	−24	24	−48
$(6)(2)(1)^2$	1	−8	21	−20	4	·	6	−30	72	−6
$(6)(2)^2$	1	−10	37	−62	44	−8	9	−48	63	−32
$(6)(3)(1)$	1	−9	27	−29	6	·	10	−60	90	
$(6)(4)$	1	−10	35	−50	26	−4				
$(7)(1)^3$	1	−7	14	−7	·	·	7	−21	7	−14
$(7)(2)(1)$	1	−9	28	−35	14	·	7	−35	112	−42
$(7)(3)$	1	−10	35	−49	21	·	10	−63	24	−48
$(8)(1)^2$	1	−8	20	−16	2	·	8	−32	88	−9
$(8)(2)$	1	−10	36	−56	34	−4	8	−48	54	−40
$(9)(1)$	1	−9	27	−30	9	−2	9	−45	100	
(10)	1	−10	35	−50	25		10	−60		

Table 5.10 (cont.)

w = 10 (11)	$a_2^5a_1^4$	$a_3^2a_2a_1^2$	$a_2^3a_1^2$	$a_1^5a_1$	$a_5a_1^2$	$a_4a_1a_1^4$	$a_4a_2^1a_1^2$	$a_4a_1^3$	$a_3a_2a_1^2$	$a_3a_2a_3a_1$
$(1)^{10}$	100800	50400	25200	16800	151200	75600	17800	15900	25200	126000
$(2)(1)^8$	-604800	-151200	201600	-151200	907200	529200	12400	17100	226800	126000
$(2)^2(1)^6$	907200	756000	554400	453600	453600	680400	453600	529200	529200	176400
$(2)^3(1)^4$	.	453600	604800	-453600	.	226800	453600	529200	226800	176400
$(2)^5$	.	.	226800	.	.	.	113400	264600 / -56700	.	113400
$(3)(1)^7$	403200	201600	100800	100800	1209600	604600	302400	151200	252000	126000
$(3)(2)(1)^5$	-1209600	-806400	504000	-604800	.	-604800	-604800	453600	907200	876600
$(3)(2)^2(1)^3$	.	604800	705600	907200	.	.	302400	453600	151200	529200
$(3)(2)^3(1)$	.	.	-302400	.	.	.	.	-151200	.	-75600
$(3)^2(1)^4$	403200 / 9	201600	100800	201600	.	.	.	.	403200	201600
$(3)^2(2)(1)^2$	9	-201600 / -18	-201600	-604800	.	.	.	.	.	-201600
$(3)^2(2)^2$	9	-36	100800 / 36	.	.	.	.	.	.	.
$(3)^3(1)$	27	-81	.	134400 / 27	.	.	.	.	.	.
$(4)(1)^6$	.	.	.	.	-907200 / -4	-453600	-226800	-113400	-151200	75600
$(4)(2)(1)^4$	.	.	.	.	-4	453600 / 8	453600	340200	453600	302400
$(4)(2)^2(1)^2$	.	.	.	.	-4	16	-226800 / -16	-340200	.	-226800
$(4)(2)^3$	.	.	.	.	-4	24	-48	113400 / 32	.	.
$(4)(3)(1)^3$	12	.	.	.	-4	12	.	.	-302400 / -12	-151200
$(4)(3)(2)(1)$	12	-24	18	36	-4	20	24	.	-12	151200 / 24
$(4)(3)^2$	33	-108	18	36	-4	24	-36	24	24	72
$(4)^2(1)^2$	16	.	.	.	-8	32	-16	.	-32	.
$(4)^2(2)$	16	-32	.	.	-8	48	-80	32	-32	64
$(5)(1)^5$	.	.	.	.	-5	.	.	.	.	.
$(5)(2)(1)^3$	.	.	.	.	-5	10	.	.	.	.
$(5)(2)^2(1)$	.	.	.	.	-5	20	-20	.	.	.
$(5)(3)(1)^2$	15	-15	.	.	-5	15	.	.	-15	.
$(5)(3)(2)$	15	-45	30	.	-5	25	-30	.	-15	30
$(5)(4)(1)$	20	-20	.	.	-9	40	-30	.	-40	20
$(5)^2$	25	-50	25	.	-10	50	-50	.	-50	50
$(6)(1)^4$	3	.	.	.	-6	6	.	.	.	.
$(6)(2)(1)^2$	3	-6	.	.	-6	18	-12	.	.	.
$(6)(2)^2$	3	-12	12	.	-6	30	-48	24	.	.
$(6)(3)(1)$	21	-45	6	9	-6	24	-18	.	-18	18
$(6)(4)$	27	-60	6	12	-10	54	-72	20	-48	72
$(7)(1)^3$	7	.	.	.	-7	14	.	.	-7	.
$(7)(2)(1)$	7	-14	.	.	-7	28	-28	.	-7	14
$(7)(3)$	28	-84	21	21	-7	35	-42	.	-28	63
$(8)(1)^2$	12	-8	.	.	-8	24	-8	.	-16	.
$(8)(2)$	12	-32	16	.	-8	40	-56	16	-16	32
$(9)(1)$	18	-27	.	3	-9	36	-27	.	-27	18
(10)	25	-60	15	10	-10	50	-60	10	-40	60

Table 5.10 (cont.)

$n-1$	$a_1a_2^4$	$a_2^2a_3^3$	$a_2^3a_3$	$a_3a_4^2$	$a_3a_2a_3$	$a_3a_2^2a_3$	$a_2a_4a_4^2$	$a_2a_3a_3$	$a_4a_4a_4$	a_5^2	$a_1a_5^2$
	4200	67200	1150	10240	151200	7560	80640	25200	1260	252	
		−75600	453600							−5040	
	201600	264600	151200							75600	
	−102400	226800	241920							50700	
	113400	56700	141750				113400				
			26350								
(3)(1)²	50400	100800	50400	604800	102400	151200		55440	15120	10080	
(3)(2)(1)²	−352800	604800	342500	604800	604800					−110880	
(3)(2)²(1)²	151200	302400	453600		302400		453600			252000	
(3)(2)²(1)	151200		151200							151200	
(3)²(1)²	151200	403200	201600							100800	
(2)²(2)(1)²	504000		201600					301600	201600	−301600	
(3)²(2)²	50400							−301600	201600	100800	
(3)³(1)								100800			
(4)(1)⁴	25200	−75600	17600	907200	453600	−226800	−151200	−75600	45360	151200	453600
(4)(2)(1)²	151200	453600	264600		453600	453600	453600	302400	151200	151200	453600
(4)(2)²(1)²	226800	−226800	151200			−226800		−226800	226800	−226800	
(4)(2)³			113400								
(4)(3)(1)²	100800	604800	−102400				102400	151200	453600	302400	102400
(4)(3)(2)(1)	302400		302400					151200	151200	102400	
(4)(3)²	−100800										
	−36									226800	226800
(4)²(1)²		226800	113400								
	16										
(4)²(2)		16	−113400	725760	162880	181440	120960	60480	10240	12096	725760
			−32	5							
(5)(1)⁶				725760	−362880	362880	362880	241920	181440	120960	
				5	−10						
(5)(2)(1)³				5	−10	181440		181440	90720	181440	
(5)(2)²(1)				5	−20	20	241920	120960	241920	241920	
							15				
(5)(3)(1)²				5	−15		−120960		241920		
							−30			
(5)(3)(2)				5	25	10	15	−181440	−362880		
								−20		
(5)(4)(1)		20		5	−20	10	20		145152		−604800
									25		
(5)²		25		10	−50	50	50	−50	−50		−6
(6)(1)⁴				6	−12						−6
(6)(2)(1)²				6	−24	24	18				−6
(6)(2)²				6	−18	12	24		−24		−6
(6)(3)(1)				6	−24						
(6)(4)	−12	24	−24								−7
(7)(1)³				7	−7	14		−21			−7
(7)(2)(1)				7	−21	21	21	8			−8
(7)(3)	−21			8	−28		8	−16			−8
(8)(1)²		4	−8	8	−16	32	18		−9		−9
(8)(2)		4		9	−32	9	30	20	−20	5	−10
(9)(1)		9	−10	10	−27	30					
(10)	−10	15		10	−40						

Table 5.10 (cont.)

$w = 10$ (iv)	$a_5a_3a_2$	$a_5a_4a_1$	$a_6a_3a_1$	a_6a_4	$a_7a_1{}^3$	$a_7a_2a_1$	a_7a_3	$a_8a_1{}^2$	a_8a_2	a_9a_1	a_{10}
$(1)^{10}$	2520	1260	840	210	720	360	120	90	45	10	1
$(2)(1)^8$	−40320	−21420	−15120	−4410	−15120	−7920	−2880	−2520	−1305	360	45
$(2)^2(1)^6$	151200	95760	75600	28980	75600	45360	20160	18900	10710	3780	630
$(2)^3(1)^4$	−151200	−151200	−126000	−69300	−75600	−75600	−50400	−37800	−28350	12600	3150
$(2)^4(1)^2$	37800	94500	37800	47250	·	37800	37800	9450	23625	9450	4725
$(2)^5$	·	−18900	·	−9450	·	·	·	·	−4725	·	945
$(3)(1)^7$	100800	50400	35280	10080	50400	25200	8640	10080	5040	1680	240
$(3)(2)(1)^5$	−403200	−252000	−226800	−100800	−302400	−176400	−80640	−100800	−55440	−25200	−5040
$(3)(2)^2(1)^3$	302400	352800	378000	252000	226800	226800	201600	151200	126000	75600	25200
$(3)(2)^3(1)$	·	−151200	25200	−100800	·	−75600	−100800	·	−75600	−25200	−25200
$(3)^2(1)^4$	100800	50400	100800	75600	201600	100800	−75600	100800	50400	33600	25200
$(3)^2(2)(1)^2$	−100800	−100800	−302400	−252000	·	·	−100800	50400	50400	−100800	−50400
$(3)^3(1)$	·	50400	67200	25200	·	·	·	50400	50400	−50400	22400
$(4)(1)^6$	−226800	−113400	−75600	−20160	−151200	−75600	−25200	−37800	−18900	−7560	−1260
$(4)(2)(1)^4$	453600	340200	302400	151200	453600	302400	151200	226800	132300	75600	18900
$(4)(2)^2(1)^2$	−226800	−340200	226800	−226800	−226800	−302400	−226800	−113400	−170100	113400	−56700
$(4)(3)(1)^3$	·	113400	−151200	75600	−302400	−151200	−100800	−302400	56700	113400	18900
$(4)(3)(2)(1)$	·	·	151200	−201600	·	151200	302400	·	−151200	−151200	−50400
$(4)^2(1)^2$	·	·	·	−50400	·	·	−100800	·	151200	151200	−50400
$(4)^2(2)$	·	·	·	113400	·	·	·	113400	56700	113400	56700
$(5)(1)^5$	362880	181440	120960	30240	362880	181440	60480	120960	60480	30240	6048
$(5)(2)(1)^3$	−362880	−362880	−362880	−181440	−362880	−362880	−241920	−362880	−241920	−181440	−60480
$(5)(2)^2(1)$	·	181440	·	−181440	·	·	181440	·	181440	−181440	60480
$(5)(3)(1)^2$	·	·	241920	90720	·	·	181440	·	181440	90720	90720
$(5)(3)(2)$	·	·	·	241920	·	·	120960	241920	120960	241920	120960
$(5)(4)(1)$	·	·	·	·	·	·	·	·	120960	·	120960
$(5)^2$	·	·	·	−181440	·	·	−120960	−120960	−120960	−181440	72576
$(6)(1)^4$	−302400	−151200	−100800	−25200	−604800	−302400	−100800	−302400	−151200	−100800	−25200
$(6)(2)(1)^2$	302400 (12)	302400	302400	151200	·	302400	302400	302400	302400	302400	151200
$(6)(2)^2$	24	−151200 (−24)	·	−75600	·	·	·	·	−151200	·	−75600
$(6)(3)(1)$	18	·	−201600 (−18)	−201600	·	·	−201600	·	·	−201600	−201600
$(6)(4)$	24	−12	−24	151200 (24)	·	·	·	·	·	·	151200
$(7)(1)^3$	·	·	·	·	518400 (7)	259200	86400	518400	259200	259200	86400
$(7)(2)(1)$	14	·	·	·	·	−259200 (−14)	−259200	·	−259200	−259200	−259200
$(7)(3)$	21	·	−21	·	·	−21	172800 (21)	·	·	·	172800
$(8)(1)^2$	8	·	·	·	·	·	·	−453600 (−8)	−226800	−453600	−226800
$(8)(2)$	24	−16	·	·	·	−16	·	·	226800 (16)	·	226800
$(9)(1)$	18	·	−9	·	9	−9	·	·	·	403200 (9)	403200
(10)	30	−10	−20	10	10	−20	10	−10	10	10	−362880 (−10)

Table 5.11

$w=11$ (i)	a_1^{11}	$a_2 a_1^9$	$a_2^2 a_1^7$	$a_2^3 a_1^6$	$a_2^4 a_1^5$	$a_2^5 a_1$	$a_3 a_1^8$	$a_3 a_2 a_1^6$	$a_3 a_2^2 a_1^4$	$a_3 a_2^3 a_1^2$	$a_3 a_2^4$
$(1)^{11}$	39916800	19958400	9979200	4989600	2494800	1247400	6652800	3326400	1663200	831600	415800
	1										
$(2)(1)^9$	1	−19958400	−19958400	−14968800	−9979200	6237000	19958400	13305600	8316000	4989600	−2910600
		−2									
$(2)^2(1)^7$	1	−4	9979200	14968800	14968800	12474000		9979200	11642400	9979200	7484400
			4								
$(2)^3(1)^5$	1	−6	13	−4989600	−9979200	−12474000			−4989600	−8316000	−9147600
				−8							
$(2)^4(1)^3$	1	−8	24	−32	2494800	6237000				2494800	5405400
					16						
$(2)^5(1)$	1	−10	40	−80	80	−1247400					−1247400
						−32					
$(3)(1)^8$	1	−3	·	·	·	·	13305600	6652800	3326400	1663200	831600
							3				
$(3)(2)(1)^6$	1	−5	6	·	·	·	3	−6652800	−6652800	−4989600	−3326400
								−6			
$(3)(2)^2(1)^4$	1	−7	16	−12	·	·	3	−12	3326400	4989600	4989600
									12		
$(3)(2)^3(1)^2$	1	−9	30	−44	24	·	3	−18	36	−1663200	−3326400
										−24	
$(3)(2)^4$	1	−11	48	−104	112	−48	3	−24	72	−96	831600
											48
$(3)^2(1)^5$	1	−6	9	·	·	·	6	−18	·	·	·
$(3)^2(2)(1)^3$	1	−8	21	−18	·	·	6	−30	36	·	·
$(3)^2(2)^2(1)$	1	−10	37	−60	36	·	6	−42	96	−72	·
$(3)^3(1)^2$	1	−9	27	−27	·	·	9	−54	81	·	·
$(3)^3(2)$	1	−11	45	81	54	·	9	72	180	−162	·
$(4)(1)^7$	1	−4	3	·	·	·	4	−8	·	·	·
$(4)(2)(1)^5$	1	−6	10	−4	·	·	4	−16	16	·	·
$(4)(2)^2(1)^3$	1	−8	22	−24	8	·	4	−24	48	−32	·
$(4)(2)^3(1)$	1	−10	38	−68	56	−16	4				
$(4)(3)(1)^4$	1	−7	14	−6	·	·	7	−24	6	·	·
$(4)(3)(2)(1)^2$	1	−9	28	−34	12	·	7	−38	54	−12	·
$(4)(3)(2)^2$	1	−11	46	−90	80	−24	7	−52	130	−120	24
$(4)(3)^2(1)$	1	−10	35	−48	18	·	10	−66	120	−36	·
$(4)^2(1)^3$	1	−8	20	−16	4	−8	8	−32	16	·	·
$(4)^2(2)(1)$	1	−10	36	−56	36	−12	8	−48	80	−32	·
$(4)^3(1)$	1	−11	44	−76	52		11	−80	172	−96	12
$(5)(1)^6$	1	−5	5	·	·	·	5	−5	·	·	·
$(5)(2)(1)^4$	1	−7	15	−10	·	·	5	−15	10	·	·
$(5)(2)^2(1)^2$	1	−9	29	−40	20	·	5	−25	40	−20	·
$(5)(2)^3$	1	−11	47	−98	100	−40	5	−35	90	−100	40
$(5)(3)(1)^3$	1	−8	20	−15	·	·	8	−35	30	·	·
$(5)(3)(2)(1)$	1	−10	36	−55	50	·	11	−51	100	−60	·
$(5)(3)^2$	1	−11	44	−75	45	·	11	−83	195	−135	·
$(5)(4)(1)^2$	1	−9	27	−30	10	·	9	−45	50	10	·
$(5)(4)(2)$	1	−11	45	−84	70	−20	9	−63	140	−110	20
$(5)^2(1)$	1	−10	35	−50	25	·	10	−60	100	−50	·
$(6)(1)^5$	1	−6	9	−2	·	·	6	−12	·	·	·
$(6)(2)(1)^3$	1	−8	21	−20	4	−8	6	−24	24	·	·
$(6)(2)^2(1)$	1	−10	37	−62	44	·	6	−36	72	−48	·
$(6)(3)(1)^2$	1	−9	27	−29	6	−12	9	−48	63	−6	·
$(6)(3)(2)$	1	−11	45	−83	64	·	9	−66	159	−132	12
$(6)(4)(1)$	1	−10	35	−50	26	−4	10	−60	50	−32	10
$(6)(5)$	1	−11	44	−77	55	−10	11	−77	165	−115	·
$(7)(1)^4$	1	−7	14	−7	·	·	7	−21	7	·	·
$(7)(2)(1)^2$	1	−9	28	−35	14	·	7	−35	49	−14	·
$(7)(2)^2$	1	−11	46	−91	84	−28	7	−49	119	−112	28
$(7)(3)(1)$	1	−10	35	−49	21	−14	10	−63	112	−42	14
$(7)(4)$	1	−11	44	−77	56	·	11	−77	161	−98	·
$(8)(1)^3$	1	−8	20	−16	2	·	8	−32	24	·	·
$(8)(2)(1)$	1	−10	36	−56	34	−4	8	−48	88	−48	6
$(8)(3)$	1	−11	44	−76	50	−6	11	−80	180	−120	·
$(9)(1)^2$	1	−9	27	−30	9	·	9	−45	54	−9	18
$(9)(2)$	1	−11	45	−84	69	−18	9	−63	144	−117	·
$(10)(1)$	1	−10	35	−50	25	−2	10	−60	100	−40	11
(11)	1	−11	44	−77	55	−11	11	−77	165	−110	·

Table 5.11 (*cont.*)

$w = 11$ (ii)	$a_1^5 a_2^3$	$a_1^3 a_2 a_3^2$	$a_1^4 a_2^2 a_3$	$a_1^2 a_2^3$	a_2^5	$a_4 a_1^7$	$a_4 a_1 a_2^3$	$a_4 a_2^2 a_3^2$	$a_4 a_2^2 a_3$	$a_4 a_2 a_3^2$	$a_4 a_1 a_3 a_4$
$(1)^{11}$	1108800	554400	277200	184800	92400	1663200	831600	415800	207900	277200	1386..
$(2)(1)^9$	−6652800	3880800	−2217600	−1663200	924000	−9979200	5821200	−3326400	1871100	−2494800	−1386800
$(2)^2(1)^7$	9979200	8316000	6098400	4989600	3326400	4989600	7484400	6652800	4989600	5821200	415800
$(2)^3(1)^5$	·	4989600	−6652800	−4989600	4989600	·	−2494800	4989600	5821200	2494800	415800
$(2)^4(1)^3$	·	·	2494800	·	2494800	·	·	1247400	3718500	·	1247400
$(2)^5(1)$	·	·	·	·	·	·	·	·	623700	·	·
$(3)(1)^8$	4435200	2217600	1108800	1108800	554400	13305600	6652800	3326400	1663200	2772000	1386000
$(3)(2)(1)^6$	−13305600	−8870400	5544000	−6652800	−3880800	·	6652800	6652800	4989600	6979200	6375600
$(3)(2)^2(1)^4$	·	6652800	7761600	9979200	8316000	·	·	6652800	4989600	1663200	5821200
$(3)(2)^3(1)^2$	·	·	−3326400	·	4989600	·	·	·	3326400	·	−831600
$(3)(2)^4$	·	·	·	·	4989600	·	·	·	·	·	·
$(3)^2(1)^5$	$\dfrac{4435200}{9}$	2217600	1108800	2217600	1108800	·	·	·	·	4435200	2217600
$(3)^2(2)(1)^3$	$\dfrac{9}{9}$	$\dfrac{-2217600}{-18}$	−2217600	−6652800	−4435200	·	·	·	·	·	−2217600
$(3)^2(2)^2(1)$	9	−36	$\dfrac{1108800}{36}$	·	3326400	·	·	·	·	·	·
$(3)^2(2)^3$	27	−81	·	$\dfrac{1478400}{27}$	739200	·	·	·	·	·	·
$(3)^3(2)$	27	−135	162	27	$\dfrac{-739200}{-54}$	·	·	·	·	·	·
$(4)(1)^7$	·	·	·	·	·	$\dfrac{-9979200}{-4}$	−4989600	−2494800	−1247400	−1663200	−831600
$(4)(2)(1)^5$	·	·	·	·	·	−4	$\dfrac{4989600}{8}$	4989600	3742200	4989600	3326400
$(4)(2)^2(1)^3$	·	·	·	·	·	−4	16	$\dfrac{-2494800}{-16}$	−3742200	·	−2494800
$(4)(2)^3(1)$	·	·	·	·	·	−4	24	−48	$\dfrac{1247400}{32}$	·	·
$(4)(3)(1)^4$	12	·	·	·	·	−4	12	·	·	$\dfrac{-3326400}{-12}$	−1663200
$(4)(3)(2)(1)^2$	12	−24	·	·	·	−4	20	−24	·	−12	$\dfrac{1663200}{24}$
$(4)(3)(2)^2$	12	−48	48	·	·	−4	28	−64	48	−12	24
$(4)(3)^2(1)$	33	−108	18	36	·	−4	24	−30	·	−24	48
$(4)^2(1)^3$	16	·	·	·	·	−8	32	−16	·	−32	72
$(4)^2(2)(1)$	16	−32	·	·	·	−8	48	−80	32	−32	64
$(4)^2(3)$	40	−144	48	48	·	−8	56	−112	48	−56	192
$(5)(1)^6$	·	·	·	·	·	−5	·	·	·	·	·
$(5)(2)(1)^4$	·	·	·	·	·	−5	10	·	·	·	·
$(5)(2)^2(1)^2$	·	·	·	·	·	−5	20	−20	·	·	·
$(5)(2)^3$	15	−15	·	·	·	−5	30	−60	40	·	·
$(5)(3)(1)^3$	15	−45	30	·	·	−5	15	−30	·	−15	·
$(5)(3)(2)(1)$	39	−165	135	45	−45	−5	25	−45	·	−30	30
$(5)(3)^2$	20	−20	40	·	·	−9	30	−30	·	−40	90
$(5)(4)(1)^2$	20	−60	·	·	·	−9	40	−110	60	−40	20
$(5)(4)(2)$	25	−50	25	·	·	−10	58	−50	·	−50	100
$(5)^2(1)$	3	·	·	·	·	−6	50	·	·	·	50
$(6)(1)^5$	3	−6	·	·	·	−6	6	−12	·	·	·
$(6)(2)(1)^3$	3	−12	12	·	·	−6	18	−48	24	·	·
$(6)(2)^2(1)$	21	−45	·	9	·	−6	30	−18	·	−18	18
$(6)(3)(1)^2$	21	−87	90	9	−18	−6	36	66	36	18	54
$(6)(3)(2)$	27	−60	6	12	·	−10	54	−72	20	−48	72
$(6)(4)(1)$	33	−105	75	15	−15	−11	66	−105	40	−60	120
$(6)(5)$	7	·	·	·	·	−7	14	·	·	−7	·
$(7)(1)^4$	7	−14	·	·	·	−7	28	−28	·	−7	14
$(7)(2)(1)^2$	7	−28	28	·	·	−7	42	−84	56	−7	28
$(7)(2)^2$	28	−84	21	21	·	−7	35	−42	·	−28	63
$(7)(3)(1)$	35	−112	42	28	·	−11	70	−126	56	−63	168
$(7)(4)$	12	−8	·	·	·	−8	24	−8	·	−16	·
$(8)(1)^3$	12	−32	16	·	·	−8	40	−56	16	−16	32
$(8)(2)(1)$	36	−140	96	36	−24	−8	48	−80	24	−40	120
$(8)(3)$	18	−27	·	3	·	−9	36	−37	·	−27	18
$(9)(1)^2$	18	−63	54	3	−6	−9	54	−99	54	−27	72
$(9)(2)$	25	−60	15	10	·	−10	50	60	10	40	60
$(10)(1)$	33	−110	60	22	−11	−11	66	110	44	55	132

Table 5.11 (*cont.*)

$\varpi=11$ (iii)	$a_6a_6a_1^5$	$a_4a_2^2a_1$	$a_3^3a_2^2$	$a_3^2a_2a_1$	$a_3^2a_1$	$a_4a_1^4$	$a_4a_2a_1^4$	$a_4a_2^3a_1^3$	$a_4a_2^1$	$a_6a_6a_1^3$	$a_6a_4a_2a_1$
$(1)^{11}$	69300	46200	69300	34650	11550	11 2640	1663 200	831160	41580	55440	27720
$(2)(1)^9$	762100	554400	831600	459450	173250	1126400	1829 600	997920	54 540	720720	388880
$(2)^2(1)^7$	2772000	2217600	2910600	1871100	900900	4989600	4158 700	2993760	1995840	2494800	1663760
$(2)^3(1)^5$	4158000	3326400	−2494800	2702700	1871100	·	2494800	1126400	1126800	2494800	2494800
$(2)^4(1)^3$	2702700	1247400	623700	1556250	1351150	·	·	1247400	2286900	·	1247400
$(2)^5(1)$	623700	·	311550	311550	·	·	·	·	623700	·	·
$(3)(1)^8$	693000	554400	1108800	554400	207900	6662500	1326400	1663200	831600	1219680	609840
$(3)(2)(1)^6$	1880800	3880800	6652800	3880800	1940400	6662500	6652800	3326400	3326400	5544000	1151840
$(3)(2)^2(1)^4$	6098400	7207200	3326400	4989600	4851000	·	4989600	4989600	4989600	4989600	5208800
$(3)(2)^3(1)^2$	3326400	1663200	·	1663200	2494800	·	1326400	1663200	1126400	·	2494800
$(3)(2)^4$	415800	·	·	·	207900	·	·	·	831600	·	·
$(3)^2(1)^5$	1108800	1663200	4435200	2217600	1108800	·	·	·	·	2217600	1108800
$(3)^2(2)(1)^3$	−2217600	−5544000	·	2217600	−4435200	·	·	·	·	2217600	2217600
$(3)^2(2)^2(1)$	1108800	554400	·	·	1108800	·	·	·	·	·	1108800
$(3)^3(1)^2$	·	1478400	·	·	1478400	·	·	·	·	·	·
$(4)(1)^7$	415800	277200	−831600	−415800	−138600	−9979200	4989600	−2494800	1247400	−1663200	831600
$(4)(2)(1)^5$	2079000	1663200	4989600	2910600	1247400	·	4989600	4989600	1742200	4989600	3326400
$(4)(2)^2(1)^3$	2910600	2494800	2494800	3742200	2910600	·	·	−2494800	−3742200	·	−2494800
$(4)(2)^3(1)$	1247400	·	·	1247400	1247400	·	·	·	1247400	·	·
$(4)(3)(1)^4$	−831600	−1108800	−6652800	3326400	−1386000	·	·	·	·	−3326400	−1663200
$(4)(3)(2)(1)^2$	1663200	3326400	·	3326400	4989600	·	·	·	·	·	1663200
$(4)(3)(2)^2$	−831600	·	·	·	−831600	·	·	·	·	·	·
	−48	·	·	·	·	·	·	·	·	·	·
$(4)(3)^2(1)$	·	−1108800	·	·	−2217600	·	·	·	·	·	·
	·	−36	·	·	·	·	·	·	·	·	·
$(4)^2(1)^3$	·	·	2494800	1247400	415800	·	·	·	·	·	·
	·	·	16	·	·	·	·	·	·	·	·
$(4)^2(2)(1)$	·	·	·	−1247400	−1247400	·	·	·	·	·	·
	·	·	16	−32	·	·	·	·	·	·	·
$(4)^2(3)$	·	·	·	·	831600	·	·	·	·	·	·
	−48	−96	16	−48	48	·	·	·	·	·	·
$(5)(1)^6$	·	·	·	·	·	7983360	3991680	1995840	997920	1330560	665280
	·	·	·	·	·	5	·	·	·	·	·
$(5)(2)(1)^4$	·	·	·	·	·	·	−3991680	−3991680	−2993760	−3991680	−2661120
	·	·	·	·	·	5	−10	·	·	·	·
$(5)(2)^2(1)^2$	·	·	·	·	·	5	·	1995840	2993760	·	1995840
	·	·	·	·	·	5	−20	20	·	·	·
$(5)(2)^3$	·	·	·	·	·	5	−30	60	−997920	·	·
	·	·	·	·	·	·	·	·	−40	·	·
$(5)(3)(1)^3$	·	·	·	·	·	5	−15	·	·	2661120	1330560
	·	·	·	·	·	·	·	·	·	15	·
$(5)(3)(2)(1)$	·	·	·	·	·	5	−25	30	·	15	−1330560
	·	·	·	·	·	5	30	45	·	30	−30
$(5)(3)^2$	·	−45	·	·	·	5	−20	10	·	20	−90
$(5)(4)(1)^2$	·	·	20	·	·	5	30	50	−20	20	−40
$(5)(4)(2)$	−40	·	20	−40	·	10	−50	50	·	50	−50
$(5)^2(1)$	·	·	25	·	·	6	·	·	·	·	·
$(6)(1)^5$	·	·	·	·	·	6	−12	·	·	·	·
$(6)(2)(1)^3$	·	·	·	·	·	6	−24	24	·	·	·
$(6)(2)^2(1)$	·	·	·	·	·	6	18	·	·	18	·
$(6)(3)(1)^2$	·	·	·	·	·	6	−30	36	·	18	−36
$(6)(3)(2)$	−36	·	·	·	·	6	−24	12	·	24	·
$(6)(4)(1)$	·	−12	24	24	·	11	−60	75	−10	60	−90
$(6)(5)$	−30	−15	30	−30	·	7	−7	·	·	·	·
$(7)(1)^4$	·	·	·	·	·	7	21	14	−28	·	·
$(7)(2)(1)^2$	·	·	·	·	·	7	−35	56	·	21	−21
$(7)(2)^2$	−28	·	·	·	·	7	−28	21	·	28	−28
$(7)(3)(1)$	·	−21	28	−56	28	7	−35	42	−14	8	·
$(7)(4)$	−42	−56	4	·	·	8	−16	·	·	·	·
$(8)(1)^3$	·	·	·	·	·	8	−32	32	·	8	−16
$(8)(2)(1)$	·	·	4	−8	·	8	−40	48	·	32	−72
$(8)(3)$	−24	−48	4	−12	12	9	27	9	·	18	·
$(9)(1)^2$	·	·	9	−18	·	9	45	63	−18	18	−36
$(9)(2)$	−36	·	9	−18	·	10	40	30	·	30	−20
$(10)(1)$	·	−10	15	−10	·	10	40	·	−11	·	−66
(11)	−33	−33	22	−33	11	11	−55	66	·	44	·

Table 5.11 (cont.)

$w=11$ (iv)	$a_2a_3^3$	$a_2a_4a_3^2$	$a_3a_4a_3$	$a_5^2a_1$	$a_4a_1^5$	$a_4a_2a_1^3$	$a_4a_3^2a_1$	$a_4a_3a_1^2$	$a_5a_2a_3$	$a_5a_4a_1$	a_4a_7
$(1)^{11}$	9240	13860	6930	2772	55440	27720	13860	9240	4620	2110	462
$(2)(1)^9$	−147840	−221760	−117810	−55440	831600	−443520	−216620	−166120	−87780	45810	11550
$(2)^2(1)^7$	776160	1081080	651420	360360	2494800	1663200	1053360	831600	498960	213820	97020
$(2)^3(1)^5$	−1663200	−1663200	1372140	−831600	−831600	1663200	−1663200	−1386000	−1108800	762760	318780
$(2)^4(1)^3$	1247400	623700	1143450	623700	.	415800	1039500	415800	900900	512760	351150
$(2)^5(1)$	.	.	−311850	.	.	.	−207900	.	−207900	−103150	103050
$(3)(1)^8$	221760	388080	194040	110880	2217600	1108800	554400	388080	194040	110880	27720
$(3)(2)(1)^6$	−1774080	−3049200	−1715640	1210680	6652800	−4435200	−2772000	−2494800	−1441440	1108800	388080
$(3)(2)^2(1)^4$	4435200	4158000	3603600	2772000	.	3326400	3880800	4158000	3326400	2772000	1386000
$(3)(2)^3(1)^2$	−3326400	−831600	−2494800	−1663200	.	.	−1663200	277200	−2217600	1108800	1386000
$(3)(2)^4$	.	.	415800	.	.	.	.	.	138600	.	138600
$(3)^2(1)^5$	776160	2217600	1108800	1108800	2217600	1108800	554400	1108800	554400	831600	388080
$(3)^2(2)(1)^3$	−3326400	2217600	−2217600	−2217600	.	−1108800	1108800	−3326400	2217600	−2772000	1663200
$(3)^2(2)^2(1)$	2772000	.	1108800	1108800	.	.	554400	.	1663200	277200	1386000
$(3)^3(1)^2$	739200	.	.	.	.	.	.	739200	369600	739200	369600
$(3)^3(2)$	−739200	.	.	.	.	.	.	.	−369600	.	369600
$(4)(1)^7$	−277200	−498960	−249480	166320	−4989600	2494800	1247400	−831600	415800	−221760	55440
$(4)(2)(1)^5$	1663200	3326400	1912680	1663200	4989600	4989600	3742200	3326400	2079000	1663200	665280
$(4)(2)^2(1)^3$	2494800	−2494800	−2910600	−2494800	.	−2494800	−3742200	−2494800	2910600	−2494800	1663200
$(4)(2)^3(1)$	.	.	1247400	.	.	.	1247400	.	1247400	831600	831600
$(4)(3)(1)^4$	1108800	−4989600	−2494800	3326400	.	.	.	−1663200	−831600	−2217600	1386000
$(4)(3)(2)(1)^2$	3326400	1663200	3326400	3326400	.	.	.	1663200	1663200	3326400	3326400
$(4)(3)(2)^2$	.	.	−831600	.	.	.	.	.	.	.	831600
$(4)(3)^2(1)$	−1108800	.	.	.	.	.	.	.	−831600	.	−554400
$(4)^2(1)^3$	.	2494800	1247400	2494800	.	.	.	.	−554400	1247400	1247400
$(4)^2(3)$	.	.	−1247400	.	.	.	.	.	−1247400	−1247400	−1247400
$(5)(1)^6$	221760	332640	166320	133056	7983360	3991680	1995840	1330560	665280	332640	77616
$(5)(2)(1)^4$	1330560	1995840	−1164240	1330560	.	−3991680	3991680	−3991680	−2661120	−1995840	831600
$(5)(2)^2(1)^2$	1995840	997920	1496880	1995840	.	.	1995840	.	1995840	997920	1496880
$(5)(2)^3$	.	.	−498960	.	.	.	.	.	.	.	−166320
$(5)(3)(1)^3$	887040	2661120	1330560	2661120	.	.	.	.	1330560	2661120	1774080
$(5)(3)(2)(1)$	−2661120	.	1330560	−2661120	.	.	.	2661120	−1330560	.	−2661120
$(5)(3)^2$	<u>887040</u> 45	.	.	.	.	.	.	.	.	.	443520
$(5)(4)(1)^2$	.	<u>−1995840</u> −20	−997920	−3991680	.	.	.	.	.	−1995840	−2993760
$(5)(4)(2)$	.	−20	<u>997920</u> 40	.	.	.	.	.	.	.	997920
$(5)^2(1)$	.	−50	.	<u>1596672</u> 25	.	.	.	.	.	.	1596672
$(6)(1)^5$	.	.	.	.	<u>−6652800</u> −6	−3326400	1663200	−1108800	−554400	277200	−5544
$(6)(2)(1)^3$	.	.	.	.	−6	<u>3326400</u> 12	3326400	3326400	2217600	1663200	554400
$(6)(2)^2(1)$	.	.	.	.	−6	12	<u>−1663200</u> −24	.	−1663200	831600	−831600
$(6)(3)(1)^2$	.	.	.	.	6	24	−24	<u>−2217600</u> −18	−1108800	−2217600	−1108800
$(6)(3)(2)$	.	.	.	.	6	18	.	−18	<u>1108800</u> 36	.	1108800
$(6)(4)(1)$	.	−24	.	.	−6	30	−36	−18	36	<u>1663200</u> 24	1663200
$(6)(5)$	.	−24	.	.	−6	24	12	−24	.	24	<u>−1330560</u> −30
$(7)(1)^4$	15	−60	30	30	−6	30	30	−30	30	30	−30
$(7)(2)(1)^2$	.	.	.	.	−7	30	.	.	.	.	.
$(7)(2)^2$	.	.	.	.	−7	14	.	.	.	.	.
$(7)(3)(1)$	.	.	.	.	−7	28	−28	.	.	.	.
$(7)(4)$	.	−28	28	.	−7	21	.	21	.	28	.
$(8)(1)^3$	.	.	.	.	−7	28	−14	28	.	.	.
$(8)(2)(1)$	24	.	.	.	−8	8	−16	.	.	.	.
$(8)(3)$	.	−9	.	.	−8	24	−24	−24	24	.	.
$(9)(1)^2$	.	−9	18	.	−8	32	−36	−9	.	24	.
$(9)(2)$	.	−20	.	5	−9	18	−10	9	.	18	.
$(10)(1)$	11	−33	22	11	−10	30	−33	20	22	10	−11
(11)	11	−33	22	11	−11	44	−33	−33	22	22	−11

Table 5.11 (cont.)

$\pi = 11$ (s)	$a_7a_1^4$	$a_7a_2a_1^2$	$a_7a_3^2$	$a_7a_4a_2$	$a_8a_1^3$	$a_9a_2^2$	$a_9a_3a_1$	a_9a_4	$a_9a_1^4$	a_9a_1	$a_{10}a_1$	a_{11}
$(1)^{11}$	7920	3960	1980	1320	110	990	495	165	110	55	11	1
$(2)(1)^9$	166320	-87120	45540	11680	5610	-27720	14355	5115	3960	2035	495	55
$(2)^2(1)^7$	831600	498960	293040	221760	77220	207900	111240	171360	135960	99990	34650	6930
$(2)^3(1)^5$	-831600	-831600	665280	554400	-263340	415800	111240	171360	101950	121275	51975	17325
$(2)^4(1)^3$	·	415800	623700	415800	111580	103950	207900	-6375	-51975	-51975	-10395	-10395
$(2)^5(1)$	·	·	-207900	·	103950	·	-51975	·	·	9240	2640	110
$(3)(1)^8$	554400	277200	138600	95040	25740	110880	55440	18480	18480	9240	2640	9240
$(3)(2)(1)^6$	-3326400	-1940400	-1108800	887040	-332640	1108800	609840	-249480	277200	147840	55440	69300
$(3)(2)^2(1)^4$	1663200	2494800	2217600	2217600	1247400	1663200	1386000	997920	831600	554400	277200	138600
$(3)(2)^3(1)^2$	·	831600	-1663200	1108800	1108800	·	-831600	-997920	277200	554400	277200	138600
$(3)(2)^4$	·	·	415800	·	207900	·	·	14850	135960	·	92400	34650
$(3)^2(1)^5$	2217600	1108800	554400	554400	277200	1108800	554400	221760	369600	184800	92400	18480
$(3)^2(2)(1)^3$	·	1108800	1108800	2217600	1663200	1108800	1108800	1108800	1108800	739200	554400	184800
$(3)^2(2)^2(1)$	·	·	554400	554400	831600	·	554400	1108800	·	554400	277200	277200
$(3)^3(1)^2$	·	·	·	739200	739200	·	·	369600	246400	123200	246400	123200
$(3)^3(2)$	·	·	·	·	·	·	·	369600	·	123200	·	123200
$(4)(1)^7$	-1663200	-831600	-415800	277200	71280	415800	-207900	69300	-415800	·	-13860	-1980
$(4)(2)(1)^5$	4989600	3326400	2079000	1663200	665280	2494800	1455300	623700	831600	467530	207900	207900
$(4)(2)^2(1)^3$	·	2494800	-2910600	2494800	1663200	1247400	1871100	1455300	1247400	1039500	623700	207900
$(4)(2)^3(1)$	·	·	1247400	·	831600	·	623700	623700	·	623700	207900	207900
$(4)(3)(1)^4$	-3326400	-1663200	-831600	-1108800	831600	3326400	1663200	693000	1663200	-831600	554400	11880
$(4)(3)(2)(1)^2$	·	1663200	1663200	1320400	1326400	·	1663200	2494800	1663200	1663200	1663200	831600
$(4)(3)(2)^2$	·	·	-831600	·	831600	·	·	415800	·	831600	·	415800
$(4)(3)^2(1)$	·	·	·	1108800	-1663200	·	·	1108800	·	·	554400	554400
$(4)^2(1)^3$	·	·	·	·	415800	1247400	623700	207900	1247400	623700	623700	207900
$(4)^2(2)(1)$	·	·	·	·	-1247400	·	-623700	623700	·	623700	-623700	-623700
$(4)^2(3)$	·	·	·	·	831600	·	·	415800	·	·	·	415800
$(5)(1)^6$	3991680	1995840	997920	665280	166320	1330560	665280	221760	332640	166320	66528	11088
$(5)(2)(1)^4$	-3991680	-3991680	2993760	2661120	1164240	-3991680	2661120	1330560	1995840	1164240	665280	-166320
$(5)(2)^2(1)^2$	·	1995840	2993760	1995840	1496880	·	1995840	1995840	997920	1496880	997920	498960
$(5)(2)^3$	·	·	-997920	·	-408960	·	·	·	-408960	·	·	-166320
$(5)(3)(1)^3$	·	·	·	1330560	1330560	2661120	1330560	887040	2661120	1330560	1330560	443520
$(5)(3)(2)(1)$	·	·	·	-1330560	-1330560	·	-1330560	-2661120	·	-1330560	-1330560	-1330560
$(5)(3)^2$	·	·	·	·	·	·	·	887040	·	·	·	443520
$(5)(4)(1)^2$	·	·	·	·	-997920	·	·	-1995840	997920	-1995840	-997920
$(5)(4)(2)$	·	·	·	·	997920	·	·	·	997920	·	798336	798336
$(5)^2(1)$	·	·	·	·	·	·	·	·	·	·	-277200	-55440
$(6)(1)^5$	-6652800	-3326400	-1663200	-1108800	-277200	-3326400	-1663200	-554400	-1108800	554400	1663200	554400
$(6)(2)(1)^3$	·	3326400	3326400	3326400	1663200	3326400	3326400	2217600	3326400	-1663200	-831600	-831600
$(6)(2)^2(1)$	·	·	-1663200	·	-831600	·	-1663200	-1663200	·	·	·	·
$(6)(3)(1)^2$	·	·	·	2217600	-2217600	·	·	-1108800	-2217600	-1108800	-2217600	-1108800
$(6)(3)(2)$	·	·	·	·	1663200	·	·	1108800	1108800	·	·	1108800
$(6)(4)(1)$	·	·	·	·	·	·	·	·	·	1663200	1663200
$(6)(5)$	·	·	·	·	·	·	·	·	·	·	-1330560
$(7)(1)^4$	5702400 / 7	2851200	1425600	950400	237600	5702400	2851200	950400	2851200	1425600	950400	237600
$(7)(2)(1)^2$	7	-2851200 / -14	-2851200	-2851200	1425600	·	-2851200	-2851200	-2851200	-2851200	-2851200	-1425600
$(7)(2)^2$	7	-28	1425600 / 28	·	712800	·	·	·	·	1425600	·	712800
$(7)(3)(1)$	7	-21	·	1900800 / 21	1900800	·	1900800	·	·	·	1900800	1900800
$(7)(4)$	7	-28	14	28	-1425600 / -28	·	·	·	·	·	·	-1425600
$(8)(1)^3$	8	·	·	·	·	-4989600 / -8	2494800	831600	-4989600	-2494800	-2494800	-831600
$(8)(2)(1)$	8	-16	·	·	·	-8	2494800 / 16	2494800	·	2494800	2494800	2494800
$(8)(3)$	8	-24	·	24	·	-8	24	-1663200 / -24	·	·	·	-1663200
$(9)(1)^2$	9	-9	·	·	·	-9	·	4435200 / 9	2217600	4435200	2217600
$(9)(2)$	9	-27	18	·	·	-9	18	·	-2217600 / -18	·	-2217600
$(10)(1)$	10	-20	·	10	·	-10	10	·	10	-3991680 / -18	-3991680
(11)	11	-33	11	22	-11	-11	22	-11	11	-11	-11	3628800 / 11

Table 5.12

$w-12$ (i)	a_1^{12}	$a_2 a_1^{10}$	$a_2^2 a_1^8$	$a_2^3 a_1^6$	$a_2^4 a_1^4$	$a_2^5 a_1^2$	a_2^6	$a_3 a_1^9$	$a_3 a_2 a_1^7$	$a_3 a_2^2 a_1^5$	$a_3 a_2^3 a_1^3$
$(1)^{12}$	479001600	239500800	119750400	59875200	29937600	14968800	7484400	79833600	39916800	19958400	9979200
	1										
$(2)(1)^{10}$	1	-239500800	-239500800	-179625600	-119750400	-74844000	44906400	239500800	159667200	99792000	-59875200
		-2									
$(2)^2(1)^8$	1	-4	119750400	179625600	179625600	149688000	112266000	.	.	119750400	137708800
			4								
$(2)^3(1)^6$	1	-6	12	-59875200	-119750400	-149688000	-149688000	.	.	59875200	99792000
				-8							
$(2)^4(1)^4$	1	-8	24	-32	29937600	74844000	112266000	.	.	.	29937600
					16						
$(2)^5(1)^2$	1	-10	40	-80	80	-14968800	-44906400	.	.	.	.
						-32					
$(2)^6$	1	-12	60	-160	240	-192	7484400	.	.	.	.
							64				
$(3)(1)^9$	1	-3	.	.	.	.	.	159667200	79833600	39916800	19958400
								3			
$(3)(2)(1)^7$	1	-5	6	.	.	.	.	3	-79833600	-79833600	-59875200
									-6		
$(3)(2)^2(1)^5$	1	-7	16	-12	.	.	.	3	-12	39916800	59875200
										12	
$(3)(2)^3(1)^3$	1	-9	30	-44	24	.	.	3	-18	36	-19958400
											-24
$(3)(2)^4(1)$	1	-11	48	-104	112	-48	.	3	-24	72	-96
$(3)^2(1)^6$	1	-6	9	.	.	.	.	6	-18	.	.
$(3)^2(2)(1)^4$	1	-8	21	-18	.	.	.	6	-30	36	.
$(3)^2(2)^2(1)^2$	1	-10	37	-60	36	.	.	6	-42	96	-72
$(3)^2(2)^3$	1	-12	57	-134	156	-72	.	6	-54	180	-264
$(3)^3(1)^3$	1	-9	27	-27	.	.	.	9	-54	81	-162
$(3)^3(2)(1)$	1	-11	45	-81	54	.	.	9	-72	189	-324
$(3)^4$	1	-12	54	-108	81	.	.	12	-108	324	-324
$(4)(1)^8$	1	-4	2	.	.	.	.	4	.	.	.
$(4)(2)(1)^6$	1	-6	10	-4	.	.	.	4	-8	.	.
$(4)(2)^2(1)^4$	1	-8	22	-24	8	.	.	4	10	16	.
$(4)(2)^3(1)^2$	1	-10	38	-68	56	-16	.	4	-24	48	-32
$(4)(2)^4$	1	-12	58	-144	192	-128	32	4	-32	96	-128
$(4)(3)(1)^5$	1	-7	14	-6	.	.	.	7	-24	6	-12
$(4)(3)(2)(1)^3$	1	-9	28	-34	12	.	.	7	-38	54	-12
$(4)(3)(2)^2(1)$	1	-11	46	-90	80	-24	.	7	-52	130	-120
$(4)(3)^2(1)^2$	1	-10	35	-48	18	.	.	10	-66	120	36
$(4)(3)^2(2)$	1	-12	55	-118	114	-36	.	10	-86	252	-276
$(4)^2(1)^4$	1	-8	20	-16	4	.	.	8	-32	16	.
$(4)^2(2)(1)^2$	1	-10	36	-56	36	-8	.	8	-48	80	-32
$(4)^2(2)^2$	1	-12	56	-128	148	-80	16	8	-64	176	-192
$(4)^2(3)(1)$	1	-11	44	-76	52	-12	.	11	-80	172	-96
$(4)^3$	1	-12	54	-112	108	-48	8	12	-96	240	-192
$(5)(1)^7$	1	-5	5	.	.	.	.	5	-5	.	.
$(5)(2)(1)^5$	1	-7	15	-10	.	.	.	5	-15	10	.
$(5)(2)^2(1)^3$	1	-9	29	-40	20	.	.	5	-25	40	-20
$(5)(2)^3(1)$	1	-11	47	-98	100	-40	.	5	-35	90	-100
$(5)(3)(1)^4$	1	-8	20	-15	.	.	.	8	-35	30	-60
$(5)(3)(2)(1)^2$	1	-10	36	-55	30	.	.	8	-51	100	-60
$(5)(3)(2)^2$	1	-12	56	-127	140	-60	.	8	-67	202	-260
$(5)(3)^2(1)$	1	-11	44	-75	45	.	.	11	-83	195	-135
$(5)(4)(1)^3$	1	-9	27	-30	10	.	.	9	-45	50	-10
$(5)(4)(2)(1)$	1	-11	45	-84	70	-20	.	9	-63	140	-110
$(5)(4)(3)$	1	-12	54	-111	100	-30	.	12	-99	266	-250
$(5)^2(1)^2$	1	-10	35	-50	25	.	.	10	-60	100	-50
$(5)^2(2)$	1	-12	55	-120	125	-50	.	10	-80	220	-250
$(6)(1)^6$	1	-6	9	-2	.	.	.	6	-12	.	.
$(6)(2)(1)^4$	1	-8	21	-20	4	.	.	6	-24	24	.
$(6)(2)^2(1)^2$	1	-10	37	-62	44	-8	.	6	-36	72	-48
$(6)(2)^3$	1	-12	57	-136	168	-96	16	6	-48	144	-192
$(6)(3)(1)^3$	1	-9	27	-29	6	.	.	9	-48	63	-6
$(6)(3)(2)(1)$	1	-11	45	-83	64	-12	.	9	-66	150	-132
$(6)(3)^2$	1	-12	54	-110	93	-18	.	12	-102	288	-282
$(6)(4)(1)^2$	1	-10	35	-50	20	-4	.	10	-60	96	-32
$(6)(4)(2)$	1	-12	55	-120	126	-56	8	10	-80	216	-224
$(6)(5)(1)$	1	-11	44	-77	55	-10	.	11	-77	165	-115
$(6)^2$	1	-12	54	-112	105	-36	4	12	-96	252	-240
$(7)(1)^5$	1	-7	14	-7	.	.	.	7	-21	7	.
$(7)(2)(1)^3$	1	-9	28	-35	14	.	.	7	-35	49	-14
$(7)(2)^2(1)$	1	-11	46	-91	84	-28	.	7	-49	118	-112
$(7)(3)(1)^2$	1	-10	35	-49	21	.	.	10	-63	112	-42
$(7)(3)(2)$	1	-12	55	-119	119	-42	.	10	-83	238	-266
$(7)(4)(1)$	1	-11	44	-77	50	-14	.	11	-77	181	-98
$(7)(5)$	1	-12	54	-112	105	-35	.	12	-96	252	-245
$(8)(1)^4$	1	-8	20	-16	2	.	.	8	-32	24	.
$(8)(2)(1)^2$	1	-10	36	-56	34	-4	.	8	-48	88	-48
$(8)(2)^2$	1	-12	56	-128	140	-72	8	8	-64	184	-224
$(8)(3)(1)$	1	-11	44	-76	50	-6	.	11	-80	180	-120
$(8)(4)$	1	-12	54	-112	106	-40	4	12	-96	248	-224
$(9)(1)^3$	1	-9	27	-30	9	.	.	9	-45	54	-9
$(9)(2)(1)$	1	-11	45	-84	69	-18	.	9	-63	144	-117
$(9)(3)$	1	-12	54	-111	99	-27	.	12	-99	270	-261
$(10)(1)^2$	1	-10	35	-50	25	-2	.	10	-60	100	-40
$(10)(2)$	1	-12	55	-120	125	-52	4	10	-80	220	-240
$(11)(1)$	1	-11	44	-77	55	-11	.	11	-77	165	-110
(12)	1	-12	54	-112	105	-36	2	12	-96	252	-240

Table 5.12 (cont.)

$w=12$ (iii)	$a_3 a_4 a_1^5$	$a_3^2 a_1^6$	$a_3^2 a_4 a_1^4$	$a_3^3 a_4^2 a_1^2$	$a_3^4 a_1^3$	$a_6^2 a_1^3$	$a_6^2 a_3 a_1$	a_9^4	$a_4 a_1^9$	$a_4 a_3 a_1^6$	$a_4 a_3^2 a_1^4$
.1²	4980600	13306600	6652800	3326400	1661200	2217600	1108800	169600	19968400	9979200	4989600
(2)(1)¹⁰	14927200	79833600	−46569600	−26611200	−14958800	−19958400	−11088000	44352800	119750400	79833600	19958400
(2)²(1)⁸	59875200	119750400	99792000	73180800	49896000	59875200	39916800	199658600	59875200	39916800	79833600
(2)³(1)⁶	109771200	.	−59875200	−79833600	−79875200	−59875200	−59875200	−19958400	.	−79833600	59875200
(2)⁴(1)⁴	14864800	.	.	29937600	54583600	.	29937600	29937600	.	.	14968800
(2)⁵(1)²	14968800	.	.	.	−14968800	.	.	.	.	.	.
(2)⁶	.	.	.	.	.	.	.	.	.	.	.
(3)(1)⁹	9979200	53222400	26611200	13305600	6652800	13306600	6652800	29916800	159667200	79833600	39916800
(1)(2)(1)⁷	39916800	159667200	−106444800	−6652800	−39916800	79833600	49896000	26611200	.	−79833600	−79833600
(3)(2)²(1)⁵	59875200	.	79833600	93139200	79833600	119750400	99792000	79833600	.	.	39916800
(3)(2)³(1)³	39916800	.	.	−39916800	−66528000	.	−59875200	−79833600	.	.	.
(3)(2)⁴(1)	9979200	.	.	.	19958400	.	.	.	.	.	.
(3)(2)⁵(1)	48	.	.	.	.	.	.	.	.	.	.
(3)²(1)⁶	.	53222400	26611200	13305600	6652800	26611200	13305600	8870400	.	.	.
	.	9	.	.	.	.	.	.	.	.	.
(3)²(2)(1)⁴	.	9	−26611200	−26611200	−19958400	−79833600	−53222400	−53222400	.	.	.
	.	.	−18	.	.	.	.	.	.	.	.
(3)²(2)²(1)²	.	9	−36	13305600	19958400	.	39916800	79833600	.	.	.
	.	.	.	36	.	.	.	.	.	.	.
(3)²(2)³	144	9	−54	108	−6652800	.	.	.	.	.	.
	.	.	.	.	−72	.	.	.	.	.	.
(3)³(1)³	.	27	−81	.	.	17740800	8870400	11827200	.	.	.
	.	.	.	.	.	27	.	.	.	.	.
(3)³(2)(1)	.	27	−135	162	.	27	−8870400	−35481600	.	.	.
	.	.	.	.	.	.	−54	.	.	.	.
(3)⁴	.	54	−324	486	.	108	324	5913600	.	.	.
	.	.	.	.	.	.	.	81	.	.	.
(4)(1)⁸	.	.	.	.	.	.	.	.	−119750400	−59875200	−29937600
	.	.	.	.	.	.	.	.	−4	.	.
(4)(2)(1)⁶	.	.	.	.	.	.	.	.	−4	59875200	59875200
	.	.	.	.	.	.	.	.	.	8	.
(4)(2)²(1)⁴	.	.	.	.	.	.	.	.	−4	16	−29937600
(4)(2)³(1)²	.	.	.	.	.	.	.	.	−4	24	−16
(4)(2)⁴	64	.	.	.	.	.	.	.	−4	32	−48
(4)(3)(1)⁵	.	12	.	.	.	.	.	.	−4	12	−96
(4)(3)(2)(1)³	.	12	−24	48	.	.	.	.	−4	20	−24
(4)(3)(2)²(1)	24	12	−48	18	.	36	.	.	−4	28	−64
(4)(3)²(1)²	.	33	108	234	−36	36	−72	.	−4	24	36
(4)(3)²(2)	72	33	−174	.	.	.	.	.	−8	32	−84
(4)²(1)⁴	.	10	.	.	.	.	.	.	.	32	−16
(4)²(2)(1)²	.	16	−32	.	.	.	.	.	−8	48	−80
(4)²(2)²	64	16	−64	64	.	.	.	.	−8	64	−176
(4)²(3)(1)	12	40	−144	48	.	48	.	.	−8	56	−112
(4)³	48	48	−192	96	.	64	.	.	−12	96	−240
(5)(1)⁷	.	.	.	.	.	.	.	.	−5	10	.
(5)(2)(1)⁵	.	.	.	.	.	.	.	.	−5	20	−20
(5)(2)²(1)³	.	.	.	.	.	.	.	.	−5	30	−60
(5)(2)³(1)	40	15	−15	.	.	.	.	.	−5	15	−30
(5)(3)(1)⁴	.	15	−45	30	.	.	.	.	−5	25	.
(5)(3)(2)(1)²	120	15	−75	120	−60	.	.	.	−5	35	−80
(5)(3)(2)²	.	39	−165	135	.	45	−45	.	−5	30	−45
(5)(3)²(1)	.	20	−20	.	.	.	.	.	−9	40	−30
(5)(4)(1)³	.	20	−60	40	.	.	.	.	−9	58	−110
(5)(4)(2)(1)	20	20	−60	210	−30	60	−60	.	−9	67	−150
(5)(4)(3)	60	47	−215	25	.	.	.	.	−10	50	−50
(5)²(1)²	.	25	−50	125	−50	.	.	.	−10	70	−150
(5)²(2)	100	25	−100	.	.	.	.	.	−6	6	.
(6)(1)⁶	.	3	−6	.	.	.	.	.	−6	18	−12
(6)(2)(1)⁴	.	3	−12	12	.	.	.	.	−6	30	−48
(6)(2)²(1)²	.	3	−18	36	−24	.	.	.	−6	42	−108
(6)(2)³	96	21	45	90	.	9	−18	.	−6	24	−18
(6)(3)(1)³	.	21	−87	324	−18	9	−162	27	−6	36	−66
(6)(3)(2)(1)	12	48	−252	6	−12	72	.	.	−6	42	−90
(6)(3)²	36	27	−60	126	−12	12	−24	.	−10	54	−72
(6)(4)(1)²	64	27	−114	75	.	12	−15	.	−10	74	−180
(6)(4)(2)	10	33	−105	198	−12	15	−72	9	−11	66	−105
(6)(5)(1)	48	42	−180	.	.	36	.	.	−12	84	−180
(6)²	.	7	.	.	.	.	.	.	−7	14	.
(7)(1)⁵	.	7	−14	.	.	.	.	.	−7	28	−28
(7)(2)(1)³	.	7	−28	28	.	.	.	.	−7	42	−84
(7)(2)²(1)	28	7	−84	21	.	21	.	.	−7	35	−42
(7)(3)(1)²	.	28	−140	189	−42	21	−42	.	−7	49	−112
(7)(3)(2)	84	28	−112	42	.	28	.	.	−11	70	−126
(7)(4)(1)	14	35	−175	175	−35	35	−35	.	−12	84	−175
(7)(5)	70	42	−8	.	.	.	.	.	−8	24	−8
(8)(1)⁴	.	12	−32	16	.	.	.	.	−8	40	−56
(8)(2)(1)²	.	12	−56	80	−32	36	−24	.	−8	56	−136
(8)(2)²	96	12	−140	96	−16	48	−32	.	−8	48	−80
(8)(3)(1)	6	36	−184	152	.	.	.	.	−12	88	−200
(8)(4)	56	44	−27	.	.	3	−6	.	−9	36	−27
(9)(1)³	.	18	−63	54	.	3	.	.	−9	54	−99
(9)(2)(1)	18	18	−210	243	−27	57	−90	9	−9	63	−135
(9)(3)	54	45	−60	15	.	10	.	.	−10	50	−60
(10)(1)²	.	25	110	135	−30	10	−20	.	−10	70	−160
(10)(2)	80	25	−110	66	.	22	−11	.	−11	66	−110
(11)(1)	11	33	−180	180	−24	40	−48	3	−12	84	−180
(12)	60	42	.	.	.	.	.	.	.	.	.

Table 5.12 (cont.)

$w = 12$ (iii)	$a_4a_2^3a_1^2$	$a_4a_2^4$	$a_4a_3a_1^3$	$a_4a_3a_2a_1^2$	$a_4a_3a_2^2a_1$	$a_4a_3^2a_1^2$	$a_4a_3^2a_2$	$a_4^2a_2^2$	$a_4^2a_2a_1^2$	$a_4^3a_1^2$	$a_4^3a_2$
$(1)^{12}$	2494800	1247400	3326400	1663200	831600	554400	277200	831600	415800	207900	138600
$(2)(1)^{10}$	22453200	12474000	29937600	16632000	9147600	6652800	3603600	9979200	5405400	2910600	2079000
$(2)^2(1)^8$	59875200	41164200	69854400	49896000	33264000	26611200	16632000	34927200	22453200	13305600	10391400
$(2)^3(1)^6$	69854400	64864800	29937600	49896000	49896000	39916800	33264000	34927200	32432400	27442800	22453200
$(2)^4(1)^4$	37422000	53638200	.	14968800	32432400	34927200	27442800	27442800	18711000	25567200	16632000
$(2)^5(1)^2$	−7484400	−22453200	.	.	−7484400	14968800	−7484400	7484400	−3742200	−11226600	3742200
$(2)^6$	.	3742200	.	.	.	.	.	.	.	1851600	.
$(3)(1)^9$	19958400	9979200	33264000	16632000	8316000	6652800	3326400	13305600	6652800	4158000	2494800
$(3)(2)(1)^7$	59875200	−39916800	−119750400	79507200	46569600	46569600	26611200	79833600	40540400	26611200	23284800
$(3)(2)^2(1)^5$	59875200	59875200	19958400	69854400	73180800	86486400	6652800	39916800	59875200	53222400	58212000
$(3)(2)^3(1)^3$	19958400	39916800	.	9979200	39916800	19958400	−53222400	.	19958400	39916800	29937600
$(3)(2)^4(1)$	.	9979200	.	.	4989600	.	9979200	.	.	9979200	24948000
$(3)^2(1)^6$	.	.	.	53222400	13305600	19958400	9979200	53222400	26611200	13305600	13305600
$(3)^2(2)(1)^4$	.	.	.	26611200	26611200	−6652800	−43243200	36590400	−26611200	26611200	53222400
$(3)^2(2)^2(1)^2$	.	.	.	.	13305600	6652800	36590400	.	.	13305600	13305600
$(1)^3(2)^3$	.	.	.	.	.	.	−3326400	.	.	.	.
$(3)^3(1)^3$	.	.	.	.	.	17740800	8870400	.	.	.	17740800
$(3)^3(2)(1)$	.	.	.	.	.	.	−8870400	.	.	.	.
$(3)^4$	.	.	.	.	.	.	.	.	.	.	.
$(4)(1)^8$	−14688000	−7484400	−19958400	9979200	−4989600	−3326400	−1663200	−9979200	−4989600	−2494800	−1663200
$(4)(2)(1)^6$	44906400	29937600	59875200	39916800	24948000	19958400	11642400	59875200	34927200	19958400	14688000
$(4)(2)^2(1)^4$	44906400	−44906400	.	29937600	34927200	−29937600	24948000	29937600	44906400	39916800	34927200
$(4)(2)^3(1)^2$	14968800	29937600	.	.	14968800	.	14968800	.	14968800	29937600	14968800
	32										
$(4)(2)^4$	.	−7484400	.	.	.	.	.	.	.	−7484400	.
	128	−64									
$(4)(3)(1)^5$	.	.	−39916800	19958400	9979200	−13305600	−6652800	79833600	39916800	19958400	16632000
			−12								
$(4)(3)(2)(1)^3$	.	.	−12	19958400	19958400	39916800	26611200	.	.	39916800	39916800	59875200
				24							
$(4)(3)(2)^2(1)$	.	.	−12	48	−9979200	.	−19958400	.	.	.	−19958400	−9979200
	48				−48						
$(4)(3)^2(1)^2$	.	.	−12	48	.	−13305600	−6652800	.	.	.	26611200
						−36					
$(4)(3)^2(2)$	.	.	−24	72	.	.	6652800	.	.	.	.
	72		−24	120	−144	−36	72				
$(4)^2(1)^4$	.	.	−32	.	.	.	.	29937600	14968800	7484400	4989600
								16			
$(4)^2(2)(1)^2$	.	.	−32	64	.	.	.	.	−14968800	−14968800	−14968800
	32		−32					16	−32		
$(4)^2(2)^2$	.	−64	−32	128	−128	.	.	.	16	7484400	.
	192									64	
$(4)^2(3)(1)$	.	.	−56	192	−48	−96	.	.	16	−48	.	9979200
	48										48
$(4)^3$	.	48	96	384	192	−192	.	.	48	−192	96	192
	192										
$(5)(1)^7$	.	.	.	.	.	.	.	.	.	.	.
$(5)(2)(1)^5$	.	.	.	.	.	.	.	.	.	.	.
$(5)(2)^2(1)^3$	40	.	.	.	.	.	.	.	.	.	.
$(5)(2)^3(1)$	.	.	.	.	.	.	.	.	.	.	.
$(5)(3)(1)^4$	.	.	−15	.	.	.	.	.	.	.	.
$(5)(3)(2)(1)^2$	.	.	−15	30	.	.	.	.	.	.	.
$(5)(3)(2)^2$	60	.	−15	60	−60	.	.	.	.	.	.
$(5)(3)^2(1)$	.	.	−30	90	.	−45	.	.	.	.	.
$(5)(4)(1)^3$	.	.	−40	20	.	.	.	20	−40	.	.
$(5)(4)(2)(1)$	60	.	−40	100	−40	.	.	20	−40	.	.
$(5)(4)(3)$	90	.	−67	260	−150	−120	60	20	−60	.	60
$(5)^2(1)^2$	.	.	50	50	.	.	.	25	.	.	.
$(5)^2(2)$	100	.	50	150	100	.	.	25	50	.	.
$(6)(1)^6$	.	.	.	.	.	.	.	.	.	.	.
$(6)(2)(1)^4$	.	.	.	.	.	.	.	.	.	.	.
$(6)(2)^2(1)^2$	24	.	.	.	.	.	.	.	.	.	.
$(6)(2)^3$	.	.	.	.	.	.	.	.	.	.	.
$(6)(3)(1)^3$	120	−48	.	.	.	.	.	.	.	.	.
$(6)(3)(2)(1)$	36	.	−18	18	.	.	.	.	.	.	.
$(6)(3)^2$	54	.	−18	54	−36	.	.	.	.	.	.
$(6)(4)(1)^2$	20	.	−36	144	−108	−54	.	54	.	.	.
$(6)(4)(2)$	164	40	−48	72	−12	.	.	24	24	−24	.
$(6)(5)(1)$	40	.	48	168	144	12	24	24	−72	48	.
$(6)^2$	132	−24	−60	120	−30	−15	.	30	−30	.	.
$(7)(2)(1)^5$	.	.	−72	216	−144	−36	36	36	−72	36	.
$(7)(2)^2(1)^3$	.	.	−7	14	.	.	.	.	.	.	.
$(7)(2)^3(1)$	.	.	−7	14	.	.	.	.	.	.	.
$(7)(3)(1)^4$	56	.	−7	28	−28	.	.	.	.	.	.
$(7)(3)(2)(1)^2$	84	.	−28	63	.	−21	.	.	.	.	.
$(7)(4)(1)$	56	.	−28	110	−120	−21	42	.	.	.	.
$(7)(5)$	105	.	63	168	42	56	.	28	−56	.	28
$(8)(1)^4$	.	.	−77	248	−140	−70	35	35	−70	.	35
$(8)(2)(1)^2$	16	.	−16	32	.	.	.	4	.	.	.
$(8)(2)^2$	128	32	−16	64	−64	.	.	4	−8	.	.
$(8)(3)(1)$	24	.	−40	120	−24	−48	.	4	−16	16	12
$(8)(4)$	144	−24	−80	288	−160	−112	32	30	−12	40	80
$(9)(1)^3$	.	.	−27	18	.	.	.	9	−112	.	.
$(9)(2)(1)$	54	.	−27	72	−36	.	.	9	−18	.	27
$(9)(3)$	81	.	−54	207	−135	−81	54	9	−27	.	.
$(10)(1)^2$	10	.	−40	60	.	−10	.	15	−10	.	.
$(10)(2)$	130	−20	−40	140	120	10	20	15	40	20	11
$(11)(1)$	44	.	−55	132	−33	−33	.	22	−33	.	36
(12)	120	−12	−72	240	−144	−72	36	30	−72	18	

Table 5.12 (cont.)

$\varpi=12$ (iv)	a_1^9	$a_2a_1^7$	$a_2a_1a_3^9$	$a_2a_1^2a_3^2$	$a_2a_1^3a_3$	$a_3a_2a_4$	$a_3a_2a_4^9$	$a_1a_2d_3^1$	$a_4a_3^2a_1^9$	$a_3a_4a_1^9$	$a_3a_4a_4$
$(1)^{12}$	34650	3991680	1995840	997920	498960	665…		2494…	110880	166320	83160
$(2)(1)^{10}$	-623700	-39916800	-21954240	-11975040	-6486480	-864…	20…	11…	-1774080	-2661120	-1413720 … 7817040
$(2)^2(1)^8$	4054050	59875200	4…	359…	3395…	20…		24…	9313920	12972960	16405680
$(2)^3(1)^6$	-11226600			-39916800	-37920960	-20937600	-200	20…	-19958400	-19958400	13721400
$(2)^4(1)^4$	12162150			14968800	27442800		140 …	7484400	14968800	7484400	-3742200
$(2)^5(1)^2$	-5613300				-7484400						
$(2)^6$	935550										
$(3)(1)^9$	831…	79833600	39916800	19958400	997920	14968800	7983360	3650040	2661120	4656960	2328480
$(3)(2)(1)^7$	-9979200	79833600	79833600	-59875200	-39916800	66528000	45…	-23950080	-21288960	-36590400	-20023680
$(3)(2)^2(1)^5$	34927200		39916800	59875200	59875200	59875200	63201600	51891840	53222400	49896000	43243200
$(3)(2)(1)$	-20937600			-19958400	-39916800		-29937600	4656960	-39916800	-9979200	-39937600 … 4980600
$(3)(2)^2(1)$	7484400				9979200			14…	9313920	26611200	13305600
$(3)^2(1)$	6652800					26611200	13305600	19958400	-39916800	-26611200	-26611200
$(3)^2(2)(1)$	-39916800					-2661120	-26611200	33204000			13305600
$(3)^3(1)$	19958400						13305600	-6652800			
$(3)^3(2)$									8870400		
$(3)^4$	17740800								8870400		
$(4)(1)^8$	-623700	-119750400	-59875200	-29937600	-14968800	-10958400	-9979200	4980600	-3326400	-5087520	-2993760
$(4)(2)(1)^6$	7484400		59875200	59875200	44906400	59875200	39916800	24948000	19958400	39916800	22952160
$(4)(2)^2(1)^4$	-26105400			-29937600	-44906400		-29937600	34927200	-29937600	-29937600	-3402… 200
$(4)(2)^3(1)^2$	2245… 200				14968800			14968800			14968800
$(4)(2)^4$	561… 300										
$(4)(3)(1)^5$	-9979200					-39916800	-19958400	9979200	-13305600	-59875200	-29937600
$(4)(3)(2)(1)^3$	59875200					19958400	19958400	39916800	19958400	3991… 800
$(4)(3)(2)^2(1)$	29937600							9979200			-9979200
$(4)(3)^2(1)^2$	39916800							-13305600			
$(4)(3)^2(2)$										29937600	14968800
$(4)^2(1)^4$	3742200										-14968800
$(4)^2(2)(1)^2$	-22453200										
$(4)^2(2)^2$	11226600										
$(4)^2(3)(1)$	29937600										
$(4)^3$	-7484400										
	-64										
$(5)(1)^7$		95800320	47900160	23950080	11975040	15966720	7983360	3991680	2661120	3991680	1995840
	5										
$(5)(2)(1)^5$	5		-47900160	-47900160	-35925120	-47900160	-31933440	19958400	-15966720	-23950080	-13070880
			-10	23950080	35925120		23950080	27941760	23950080	11975040	17962560
$(5)(2)^2(1)^3$	5		-20	20	-11975040			11975040			-5987520
$(5)(2)^3(1)$	5		-30	60	-40	31933440	15966720	7983360	10644480	31933440	15966720
$(5)(3)(1)^4$	5		-15	.		15	-15966720	-15966720	-31933440	.	-15966720
$(5)(3)(2)(1)^2$	5		-25	30		15	-30	7983360	.	.	.
$(5)(3)(2)^2$	5		-35	80	-60	15	-60	60	10644480	-23950080	-11975040
$(5)(3)^2(1)$	5		-30	45		30	-90		45	-20	
$(5)(4)(1)^3$	5		-20	10		20					11975040
$(5)(4)(2)(1)$	5		-30	50	-20	20	-40	30	60	-20	40
$(5)(4)(3)$	5		-35	70	-30	35	-120			-20	60
$(5)^2(1)^2$	5		-50	50		50	-50	100		-50	100
$(5)^2(2)$	6		-70	150	-100	90	-150				
$(6)(2)(1)^4$	6		-12								
$(6)(2)^2(1)^2$	6		-24	24							
$(6)(2)^3$	6		-36	72	-48	18					
$(6)(3)(1)^3$	6		-18	36		18	-36			-24	
$(6)(3)(2)(1)$	6		-30	54		36	-108		54	-24	48
$(6)(3)^2$	6		-36	12		24	-48			-60	30
$(6)(4)(1)^2$	6		-24	60	-24	24	-90		15	-72	72
$(6)(4)(2)$	11		-30	75	-10	60	-144		36		
$(6)(5)(1)$	12		-60	108	-24	72					
$(6)^2$	7		-72	14							
$(7)(2)(1)^3$	7		-7								
$(7)(2)^2(1)$	7		-21	56	-28	21	-21	42			28
$(7)(3)(1)^2$	7		-35	21		21	-63			-28	105
$(7)(3)(2)$	7		-28	77	-42	28	-28	70	35	-70	
$(7)(4)(1)$	12		-42	42	-14	70	-175				
$(7)(5)$	8		-35	140	-70	8	-16				
$(8)(2)(1)^2$	8		-77	32		8	-32	32			64
$(8)(2)^2$	8		-16	96	-64	32	-72		24	-32	
$(8)(3)(1)$	8		-32	48	-32	40	-96	16	32		
$(8)(4)$	-16		-48	80						-9	18
$(9)(1)^3$			-40	9	-18	18	-36			-9	27
$(9)(2)(1)$			-27	63	-27	45	-135	27	54	-9	
$(9)(3)$			-45	90		30	-20			-20	40
$(10)(1)^2$			-54	30	-60	30	-80	40	11	-31	22
$(10)(2)$			-40	110	-11	44	-66	36	36	-48	72
$(11)(1)$	-4		-55	66	-48	60	-144				
(12)			-72	120							

Table 5.12 (cont.)

$w - 12$ (v)	$a_4 a_4 a_1$	$a_4{}^2 a_1{}^2$	$a_4{}^3 a_1$	$a_4 a_1{}^6$	$a_4 a_4 a_1{}^4$	$a_4 a_4{}^2 a_1{}^2$	$a_4 a_1{}^3$	$a_4 a_4 a_1{}^3$	$a_4 a_4 a_4 a_1$	$a_5 a_4{}^2$	$a_4 a_4 a_1{}^3$
$(1)^{11}$	27720	33264	16632	665280	332640	166320	83160	110880	55440	18480	27720
$(2)(1)^{10}$	−526680	−665280	−349272	−9979200	−5322240	−2827440	−1496880	−1995840	−1053360	−388080	−582210
$(2)^2(1)^8$	3492720	4324320	2494800	29937600	19958400	12640320	7733880	9979200	5987520	2661120	3820...
$(2)^3(1)^6$	−9812880	−9979200	−7151760	−9979200	−19958400	−19958400	−16299360	−16632000	−13305600	−7761600	914...
$(2)^4(1)^4$	11226600	7484400	8731800	.	−19958400	12474000	16210200	4989600	10810800	9147...	6217...
$(2)^5(1)^2$	−3742200	.	−3742200	.	4989600	12474000	−7484400	.	−2494800	−2494800	124740...
$(2)^6$	.	.	.	.	.	.	1247400	.	.	.	.
$(3)(1)^9$	831600	1330560	665280	26611200	13305600	6652800	3326400	4656960	2328480	813120	110800
$(3)(2)(1)^7$	−9313920	−14636160	−7983360	−79833600	−53222400	−33264000	−19958400	−29937600	−17297280	708...	111.46...
$(3)(2)^2(1)^5$	30035520	33264000	23950080	.	39916800	46560960	39916800	49896000	39916800	26611200	33264000
$(3)(2)^3(1)^3$	−33264000	−19958400	−26611200	.	.	−19958400	−33264000	−3326400	−26611200	−31046400	13305600
$(3)(2)^4(1)$	7484400	.	9979200	.	.	.	9079200	.	1663200	332640...	.
$(3)^2(1)^6$	5987520	13305600	6652800	26611200	.	13305600	6652800	3326400	6652800	3769920	9979200
$(3)^2(2)(1)^4$	−29937600	−26611200	−19958400	.	−13305600	−13305600	−9979200	−39916800	−26611200	−23284800	332640000
$(3)^2(2)^2(1)^2$	29937600	13305600	19958400	.	.	6652800	.	9979200	19958400	36590400	3326400
$(3)^2(2)^3$	−3326400	.	−6652800	.	.	.	−3326400	.	.	−1108800	.
$(3)^3(1)^3$	8870400	.	.	.	.	.	.	8870400	4435200	5913600	8870400
$(3)^3(2)(1)$	−8870400	.	.	.	.	.	.	.	−4435200	−17740800	.
										2956800	
$(4)(1)^8$	−997920	−1995840	−997920	−59875200	−29937600	−14968800	−7484400	−9979200	−9989600	−1663200	−2661120
$(4)(2)(1)^6$	9646560	19958400	10077120	59875200	59875200	44906400	29937600	39916800	24948000	11642400	19958400
$(4)(2)^2(1)^4$	−24948000	−29937600	−24948000	.	.	−29937600	−44906400	−44906400	−29937600	−34927200	−24948000	29937600
$(4)(2)^3(1)^2$	14968800	.	14968800	.	.	.	14968800	29937600	.	14968800	14968800	9979200
$(4)(2)^4$	.	.	.	.	.	.	.	−7484400	.	.	.	.
$(4)(3)(1)^5$	−11975040	−39916800	−19958400	.	.	.	.	.	−19958400	−9979200	−6652800	−26611200
$(4)(3)(2)(1)^3$	46569600	39916800	39916800	.	.	.	.	.	19958400	19958400	26611200	39916800
$(4)(3)(2)^2(1)$	19958400	.	−19958400	.	.	.	.	.	.	−9979200	−19958400	.
$(4)(3)^2(1)^2$	−19958400	.	.	.	.	.	.	.	.	.	−6652800	−6652800
$(4)(3)^2(2)$	6652800	.	.	.	.	.	.	.	.	.	6652800	.
$(4)^2(1)^4$	4989600	27937600	14968800	.	.	.	.	.	.	.	.	14968800
$(4)^2(2)(1)^2$	−14968800	.	−14968800	.	.	.	.	.	.	.	.	−14968800
$(4)^2(2)^2$	.	.	.	.	.	.	.	.	.	.	.	.
$(4)^2(3)(1)$	9979200	.	.	.	.	.	.	.	.	.	.	.
$(4)^3$	.	.	.	.	.	.	.	.	.	.	.	.
$(5)(1)^7$	665280	1596672	798336	95800320	47900160	23950080	11975040	15966720	798336	2661120	3991680
$(5)(2)(1)^5$	−598752	−15966720	−8781696	.	−47900160	−47900160	−35925120	−47900160	−31933440	−15966720	−23950080
$(5)(2)^2(1)^3$	1970880	23950080	1958400	.	.	23950080	35925120	.	23950080	23950080	11975040
$(5)(2)^3(1)$	−598752	.	−11975040	.	.	.	−11975040	.	.	.	.
$(5)(3)(1)^4$	6652800	31933440	15966720	.	.	.	.	.	31933440	15966720	10644480	31933440
$(5)(3)(2)(1)^2$	−23950080	−31933440	−31933440	.	.	.	.	.	.	−15966720	−31933440	.
$(5)(3)(2)^2$	3991680	.	15966720	.	.	.	.	.	.	.	.	.
$(5)(3)^2(1)$	10644480	.	.	.	.	.	.	.	.	.	10644480	.
$(5)(4)(1)^3$	−3991680	−47900160	−23950080	.	.	.	.	.	.	.	.	−23950080
$(5)(4)(2)(1)$	11975040	.	23950080	.	.	.	.	.	.	.	.	.
$(5)(4)(3)$	−7983360	.	.	.	.	.	.	.	.	.	.	.
$(5)^2(1)^2$	$\dfrac{-60}{}$ 19160064	9580032	.	.	.	.	.	.	.	.	.	
	25											
$(5)^2(2)$	$\dfrac{-9580032}{25}$	-50	.	.	.	.	.	.	.	.	.	
$(6)(1)^6$	$\dfrac{-79833600}{-6}$	−39916800	−19958400	−9979200	−13305600	−6652800	−2217600	−3326400				
$(6)(2)(1)^4$	−6	$\dfrac{39916800}{12}$	39916800	39916800	29937600	39916800	26611200	13305600	19958400			
$(6)(2)^2(1)^2$	−6	24	$\dfrac{-19958400}{-24}$	−29937600	.	−19958400	−19958400	−9979200				
$(6)(2)^3$	−6	36	−72	$\dfrac{9979200}{48}$	.	.	.	.				
$(6)(3)(1)^3$	−6	18	.	$\dfrac{-26611200}{-18}$	−13305600	−8870400	−26611200					
$(6)(3)(2)(1)$	−6	30	−36	.	$\dfrac{13305600}{36}$	26611200						
$(6)(3)^2$	−6	36	−54	.	−18	$\dfrac{-8870400}{-54}$						
$(6)(4)(1)^2$	−6	24	−12	−36	108	19958400	24					
$(6)(4)(2)$	−6	36	−60	−24	−24	48	24					
$(6)(5)(1)$	30	−6	30	−30	−30	30	30					
$(6)^2$	36	−12	72	−108	24	−72	144	−36	72			
$(7)(1)^5$	−7	14	.	.	.	.	.	.	.			
$(7)(2)(1)^3$	−7	28	−28	.	.	.	.	.				
$(7)(2)^2(1)$	−7	21	.	.	.	−21	.	.				
$(7)(3)(2)$	−7	35	−42	.	.	−21	42	.				
$(7)(4)(1)$	−7	28	−14	.	.	−28	.	28				
$(7)(5)$	−35	35	−35	−7	35	−35	−35	35	35			
$(8)(1)^4$	−8	8	.	.	.	.	.					
$(8)(2)(1)^2$	−8	24	−16	.	.	.	.					
$(8)(2)^2$	−8	40	−64	32	.	−24	24	.				
$(8)(3)(1)$	−8	32	−24	.	.	−24	24	.				
$(8)(4)$	−32	−8	40	−48	16	−32	32	32				
$(9)(1)^3$	−9	18	.	.	.	.	.					
$(9)(2)(1)$	−9	36	−36	.	−9	18	.					
$(9)(3)$	−27	−9	45	−54	−9	−36	81	−27				
$(10)(1)^2$	5	−10	30	−10	−10	−20	.	10				
$(10)(2)$	5	−10	50	−70	20	−20	40	.	10			
$(11)(1)$	11	−10	44	−33	.	−33	22	.	22			
(12)	−24	18	−12	−12	60	−72	12	−48	72	−12	36	

Table 5.12 (cont.)

$\kappa = 12$ (XII)	$a_1 a_2 a_3$	$a_4 a_2 a_1$	a_2^4	$a_5 a_2^2$	$a_4 a_3 a_2^2$	$a_5 a_3^2 a_2$	$a_6 a_3 a_2^2$	$a_6 a_4 a_2$	$a_6 a_5$	$a_7 a_5^2$
(1)...	13860	5544	924	95040	47520	23760	11880	7920	1980	792	11880
(6)(4)(2)	-9979200										
	-48										
(6)(5)(1)		-15966720	-31933440								
		-30									
(6)²			13305600								
			36								
	-72	-72		68428800	34214400	17107200	11404800	5702400	2851200	570240	68428800
(7)(1)⁵				7							
					-34214400	-34214400	-34214400	-22809600	-17107200	-5702400	
(7)(2)(1)³				7	-14						
					17107200		17107200	8553600	8553600		
(7)(2)²(1)				7	-28	28	22809600	11404800	22809600	11404800	
(7)(3)(1)²				7	-21		21	-11404800		-11404800	
								-42			
(7)(3)(2)				7	-35	42	21		-17107200	-17107200	
									-28		
(7)(4)(1)				7	-28	14	28		13685760		
									35		
(7)(5)				7	-35	35	35	-35	-35	-59875200	
(8)(1)⁴		-35		8							-8
				8	-16						-8
(8)(2)(1)²				8	-32	32					-8
(8)(2)²				8	-24	24	24				-8
(8)(3)(1)	-32			8	-32	16	32		-32		
(8)(4)											-9
(9)(1)³				9	-9	18					-9
(9)(2)(1)				9	-27	27	27	-27			-10
(9)(3)				9	-36		10				-10
(10)(1)²				10	-20		10	-20			-11
(10)(2)	-20			10	-40	40	40		-11		-12
(11)(1)		11		11	-33	11	22	-24	-24	12	
(12)	-24	24	6	12	-48	36	36				

Table 5.12 (cont.)

$w-12$ (vii)	$a_6a_5a_1^2$	$a_6a_4^3$	$a_6a_5a_3$	a_5a_4	$a_5a_3^3$	$a_5a_4a_1$	a_6a_6	$a_{12}a_3^2$	$a_{12}a_2$	$a_{11}a_1$	a_{12}
(1)¹²	5940	2070	1980	495	1320	660	220	132	66	12	
(2)(1)¹⁰	172260	89100	61380	16830	47520	24120	8880	5940	1056	66	
(2)²(1)⁸	1413720	792900	582120	188595	498960	273240	106920	81180	14550	1188	
(2)³(1)⁶	3742200	2577900	2079000	871180	1663200	1051080	528080	415800	249480		
(2)⁴(1)⁴	3118500	3430350	2702700	1611225	1247400	1455300	1039500	623700	519750		
(2)⁵	623700	1871100	623700	935550		623700	623700	124740	374220	12474	6237
(2)⁶		311850		155925					62370		10395
(3)(1)⁹	665280	332640	225720	50400	221760	110880	37400	31680	15840	792	44
(3)(2)(1)⁷	−7318080	3991680	2993760	997920	3326400	1774800	681120	665280	348480		
(3)(2)²(1)⁵	16632000	11975040	10810800	5155920	9979200	6652800	3492720	3326400	1995840		
(3)(2)³(1)	−9979200	−13305600	−11642400	8316000	3326400	−6652800	6652800	3326400	3326400		
(3)(2)⁴(1)		4989600	415800	2910600		1663200	2079000				
(3)²(1)⁶	6652800	3326400	2661120	997920	4435200	2217600	831120	1108800	554400		
(3)²(2)(1)⁴	−13305600	9979200	13305600	8316000	−13305600	8870400	5544000	−6652800			
(3)²(2)²	6652800	9979200	13305600	11642400		6652800	9979200	3326400	4989600	3326400	1663200
(3)³(2)		−3326400		−1663200			1108800				
(3)³(2)(1)			4435200	4435200	2956800	1478400	1971200	2956800			
(3)⁴			−4435200	4435200		1478400	5913600				
(4)(1)⁸	−2494800	−1247400	−831600	−210870	−997920	−498960	−166320	−166320	83160	2376	2970
(4)(2)(1)⁶	17463600	9979200	7484400	2577960	9979200	5488560	2162160	2494800	1330560		8316
(4)(2)²(1)⁴	22453200	−19958400	−17463600	9355500	−14968800	−12474000	7484400	7484400	4989600	2494800	623700
(4)(2)³(1)²	7484400	14968800	7484400	8731800		7484400	7484400	2494800			1247400
(4)(2)⁴		−3742200		−2182950					−1247400		−311850
(4)(3)(1)⁵	−19958400	9979200	8316000	3659640	−19958400	−9979200	3659040	6652800	3326400		332640
(4)(3)(2)(1)³	19958400	1958400	29937600	21284800	19958400	19958400	16612000	19958400	13305600		
(4)(3)(2)²(1)		−9979200	4989600	14968800		9979200	14968800				
(4)(3)²(1)²			−13305600	−16612000			6652800	6652800			
(4)(3)²(2)				3326400			6652800				
(4)²(1)⁴	7484400	3742200	2494800	1871100	14968800	7484400	2494800	7484400	3742200	24948	623700
(4)²(2)(1)²	7484400	−7484400	7484400	−11226600		7484400	7484400	7484400	7484400	7484400	3742200
(4)²(2)²		3742200		5613300					3742200		
(4)³(3)(1)			4989600	14968800			4989600			4989600	
(4)⁴				−3742200							
(5)(1)⁷	7983360	3991680	2661120	665280	3991680	1995840	665280	798336	399168		19008
(5)(2)(1)⁵	31933440	−19958400	15966720	5987520	23950080	11970880	5987520	−7983360	4390848		
(5)(2)²(1)³	23950080	27941760	23950080	13970880	11975040	17962560	13970880	11975040	9979200		
(5)(2)³(1)		−11975040		−5987520		−5987520	−5987520		−5987520	−11975040	
(5)(3)(1)⁴	15966720	7983360	10644480	6652800	31933440	15966720	6652800	15966720	7983360		
(5)(3)(2)(1)²	−15966720	−15966720	31933440	23950080		15966720	23950080	15966720	15966720	15966720	7983360
(5)(3)(2)²		7983360		3991680			3991680		7983360		3991680
(5)(3)²(1)			10644480	10644480			10644480			5222240	5222240
(5)(4)(1)³				3991680	−23950080	11975040	−3991680	23950080	−11975040	−11975040	3991680
(5)(4)(2)(1)				11975040		11975040	11975040	11975040	11975040	11975040	11975040
(5)(4)(3)				−7983360			7983360				7983360
(5)²(1)²								9580032	4790016	9580032	4790016
(5)²(2)								−4790016			−4790016
(6)(1)⁶	−19958400	−9979200	−6652800	−1663200	−13305600	−6652800	−2217600	−3326400	1663200	6652800	110880
(6)(2)(1)⁴	39916800	29937600	26611200	11642400	39916800	26611200	13305600	19958400	11642400	6652800	1663200
(6)(2)²(1)²	−19958400	−29937600	−19958400	−14968800		−19958400	−19958400	9979200	14968800	9979200	4986800
(6)(2)³		9979200		4989600			−8870400	−26611200	4989600		1663200
(6)(3)(2)(1)			−13305600	−13305600	−26611200	−13305600	26611200		−13305600	−13305600	−4435200
(6)(3)(1)³			13305600	13305600		13305600	−8870400				1310600
(6)(3)²											−443520
(6)(4)(1)²				9979200				19958400	9979200	19958400	9979200
(6)(4)(2)				−9979200				−9979200			9979200
(6)(5)(1)										−15966720	−15966720
(6)²											6652800
(7)(1)⁵	34214400	17107200	11404800	2851200	34214400	17107200	5702400	11404800	5702400	2851200	570240
(7)(2)(1)³	34214400	−34214400	−34214400	17107200	34214400	34214400	22809600	−34214400	−22809600	17107200	5702400
(7)(2)²(1)		17107200		8553600			17107200	17107200		17107200	8553600	8553600
(7)(3)(1)²			22809600	22809600			11404800	22809600	11404800	22809600	11404800
(7)(3)(2)							−11404800		−11404800		11404800
(7)(4)(1)				17107200						17107200	17107200
(7)(5)											13685760
(8)(1)⁴	−29937600	−14968800	−9979200	−2494800	−59875200	−29937600	−9979200	−29937600	−14968800	−9979200	2494800
(8)(2)(1)² 16	29937600	29937600	29937600	14968800		29937600	29937600	29937600	29937600	29937600	14968800
(8)(2)² 32		−14968800 −32		−7484400					−14968800		−7484400
(8)(3)(1) 24			−19958400 −24	−19958400			−19958400				−19958400 −19958400
(8)(4) 32				14968800 32							149688+..
(9)(1)³					53222400 9	26611200	8870400	53222400	26611200	26611200	8870400
(9)(2)(1) 18					9	−26611200 −18	26611200		26611200	26611200	26611200
(9)(3) 27			−27		9	−27	17740800 27				17740800
(10)(1)² 10					10		−47900160 −10	−23950080		−47900160	−23950080
(10)(2) 30	−20			10	−20		−10	23950080 20		23950080
(11)(1) 22		−11		11	−11		−11		43545600 11	43545600
(12) 36	−12	−24	12	12	−24	12	−12	12	12	−39916800 −1..

CONTROL CHARTS WITH WARNING LINES

By E. S. PAGE

Statistical Laboratory, University of Cambridge[*]

Process inspection schemes using control charts with warning lines are considered and the properties of some schemes based on the observations from the last few samples are evaluated. Tables of schemes for controlling the mean of a normal population are given. A method is suggested for controlling both the mean and standard deviation of a population on a single chart; examples are given when the population distribution is normal.

1. Introduction

A problem that arises in industrial applications of statistics is to detect changes in the para-meters specifying the quality of the output from a continuous production process, so that some rectifying action can be taken to restore the parameters to satisfactory values. A widely used scheme for this purpose (a *process inspection scheme*) is based on a control chart (Shewhart, 1931). Samples of fixed size are taken at regular intervals and a statistic of the sample (for example, the mean, the range, or the number of defectives) is plotted on the chart; if the sample point falls outside control limit(s) drawn on the chart, i.e. if the statistic differs by more than a given amount from its satisfactory value, rectifying action is taken. This scheme will be referred to as a *single-sample* scheme, since the decision whether or not to take action is based upon a single point on the chart; also in what follows the lines on the chart denoting a serious departure of the statistic from its satisfactory value will be called *action lines* instead of control limits. With any process inspection scheme there will in general be a delay between the point at which the parameters changed and that at which the scheme demands action. During this delay the production fails to meet the quality requirements. On the other hand, in practical cases, process inspection schemes, in particular single-sample schemes, cause action to be taken even when the parameters constantly maintain satis-factory values. In the one case there is a loss due to the production of substandard material, and in the other due to unnecessary interference with the production process. These losses, as a function of the parameters, may be used as a basis for the selection of process inspection schemes or for the comparison of two schemes. When the fraction of the production that is inspected is constant, a convenient assessment of these losses can be obtained from the average number of articles that are inspected before the scheme requires rectifying action when the parameters remain constant: this number, a function of the parameter values, has been called the *average run length* (Page, 1954a; cf. Aroian & Levene, 1950). We shall base the choice of inspection schemes on the behaviour of their average run length functions.

Consider a single-sample control chart scheme for controlling the mean of a normal distribution. Suppose that the standard deviation, σ, of the process is known, and that the 'ideal' value for the process mean is μ. Then samples of size N are examined at regular in-tervals, and if the mean, $\bar{x}$, falls outside the action lines drawn at $\mu \pm B_1 \sigma / \sqrt{N}$, rectifying action is taken. The current practice is to choose N, usually small, and to use 'three-sigma'

* Present address: University of Durham.

limits, i.e. to take $B_1 = 3$. These action lines are chosen so that the chance is about 0·002 that a given sample point will cause action to be taken when the mean is at its satisfactory value μ. This consideration gives no guidance about the best sample size to choose nor any information about the average amount of time that will elapse before a change in the mean is noticed. If the fraction of production that may be inspected is fixed, for example, by a limitation on the man-hours for inspection, it is more realistic to choose the scheme with the most suitable average run length function. Tables giving values of N and B_1 have been constructed to enable this choice to be made easily for controlling the mean of a normal distribution (Page, 1954*b*), and similar tables could be computed for other situations. It turns out that the best sample size to use is often much larger than those customarily taken in industrial practice, so that samples should be taken less frequently. But although the theoretically best sample size is large it may not be practically convenient to take samples of this size; further, the quality control engineer has got used to small samples and he may be reluctant to change his habits radically. There is, too, his (correct) feeling that a small sample will spot a very serious change in the mean, and for this reason he will wish to continue taking small samples frequently. It is therefore of interest to seek schemes of the control chart type that are easy to apply, that require only small samples to be taken, and yet retain the advantages of the single-sample schemes using large samples. One method is suggested by a modification of the single-sample scheme that has occasionally been used (Dudding & Jennett, 1944); in this a cluster of 'moderately' extreme sample points is treated as a single point outside the action lines and accordingly action is taken. A point is adjudged 'moderately' extreme if it lies between *warning lines* and the action lines.* Thus for controlling the mean of a normal population with known standard deviation, σ, at an ideal value μ warning lines would be drawn at $\mu \pm B_2 \sigma / \sqrt{N}$, where $B_2 < B_1$.

We shall consider rules of the following type:

I. *Choose k, n, N. Take samples of size N. Take action if any point falls outside the action lines or if any k out of the last n points fall outside the warning lines.*

In the next section we show that in a certain sense it is reasonable to consider only two special cases of rule I, and in the following section their properties are evaluated.

2. Restriction of the rules

In the rules of type I only the region of the chart in which a sample point falls is taken into account when deciding whether or not to take action; the position of the sample point within, say, the warning region is not considered. For the rules it is only necessary to count the number of points in the various regions; however, for mathematical convenience it is useful to suppose that a score is assigned to each point according to the region in which it falls. A point outside the warning or action lines is obtained when the quality of the sample falls below the required level; if a mark or score is to be assigned to the sample such a defection will merit some penalty. On the other hand, a point within the warning lines will receive a bonus score. It is reasonable to base a process inspection scheme on these scores; an accumulation of penalties will be taken to indicate a deterioration in quality and suitable rectifying action will be taken. After each sample is taken the total penalty score over the last 'few' samples is examined and action is taken if it is 'large'. In order to make the scheme precise, let a score x_i be assigned to the ith sample, where $x_i = -a, b, c$ $(a > 0, c > b > 0)$.

* A chart using only warning lines has been considered by Weiler (1953).

according as the sample point falls within the warning lines, between the warning and action lines, or outside the action lines, respectively. The scheme is:

11. *Take action after the n-th sample if any of the inequalities*

$$\sum_{i=0}^{r} x_{n-i} > h_{r+1} \quad (r = 0, 1, ..., s)$$

are satisfied, where s and the h_r are suitably chosen constants.

We consider the case where s is limited only by the number of samples drawn since action was last taken, so that all the partial sums, working back from the last sample, are examined: and where $h_r = h$, all r, so that the total scores in the last one, two, three, ..., samples are tested to see that none is greater than h. In an earlier paper (Page, 1954 a) it was shown that this scheme is equivalent to a sequence of linear sequential tests (Barnard, 1946; Wald, 1947) with initial score on the acceptance boundary, if the test ends on the acceptance boundary the test is reapplied, while if it ends on the rejection boundary action is taken. Consequently we consider what restrictions must be placed on k and n in order that the scheme I shall be equivalent to the repeated application of a linear sequential test with boundaries at $(0, h)$ and initial score zero.*

First, we have that an initial sequence of $(k-1)$ points between warning and action lines has total penalty score $(k-1)b$; since this sequence does not require action to be taken

$$(k-1)b < h. \tag{1}$$

If now this sequence is followed by $(n-k)$ points within the warning lines each receiving a bonus score $-a$, and then by another point between the warning and action lines, scheme I requires action to be taken. Hence the total penalty score must be at least h, i.e.

$$kb - (n-k)a \geqslant h. \tag{2}$$

Further, any sequence of $(n-k+1)$ bonus points (i.e. between the warning lines) is equivalent to restarting the scheme: in particular, the set consisting of $(k-1)$ points between the warning and action lines followed by $(n-k+1)$ bonus points must have a total penalty score at most zero. Hence $\quad (k-1)b - (n-k+1)a \leqslant 0:$ $\tag{3}$

indeed, if this inequality were not satisfied a finite number of such sets of n points would cause the total penalty score to exceed h and action to be taken, contrary to the conditions I. The combination of (1) and (2) gives $\quad b/a > n - k,$ $\tag{4}$

and hence with (3) we obtain $\quad (n-k) < (n-k+1)/(k-1),$ $\tag{5}$

since $k \geqslant 2$ for the warning lines to be distinct from the action lines. It follows that

$$n < \frac{(k-1)^2}{k-2} = k + \frac{1}{k-2}. \tag{6}$$

For scheme I, k and n are integral and $2 \leqslant k \leqslant n$, so that (6) can be satisfied only for

(i) $k = 2$, any n,

or (ii) $k = n$.

* The restriction on the initial score is unnecessary; if the initial score is Z $(0 \leqslant Z < h)$, inequalities are obtained in a similar way and found to imply $Z = 0$.

The two rules we consider therefore are:

III. *Choose n, N. Take samples of size N. Take action if two points out of any sequence of n fall between the warning and action lines or if any point falls outside the action lines.*

IV. *Choose n, N. Take samples of size N. Take action if n consecutive points fall between the warning and action lines or if any point falls outside the action lines.*

In these two cases it cannot happen that some sets of points not requiring action to be taken have greater total penalty scores than some sets leading to action; for any other values of k, n the anomalous position can, however, occur. The above theory admits the corollary that no other types of warning lines, more or less extreme than those considered above, may be introduced and yet permit the equivalence of a rule of type I based on the last n points, and a rule of type II. In the next section we derive the average run lengths of the schemes III and IV.

3. THE AVERAGE RUN LENGTHS OF THE RULES

Although we shall not use the following method, for completeness we remark that the average run length of the general rule of type I may be evaluated by enumerating the possible combinations of the last $(n-1)$ points on the chart such that action has not been required, and treating the combinations as the states of a discrete Markov chain (e.g. Bartlett, 1953; Feller, 1950). The matrix of transition probabilities, $\mathbf{P}$, may be written down; the vector giving the probability that the chart is in a given state after the rth point is $\mathbf{p}_r = \mathbf{P}^r \mathbf{p}_0$, where $\mathbf{p}_0$ is the vector specifying the initial state. Accordingly, the probability that action has not been taken up to and including the rth sample is the sum of the elements of $\mathbf{p}_r$, i.e. $\mathbf{1}'\mathbf{p}_r$, where $\mathbf{1}' = (1, 1, \ldots, 1)$. Hence the probability that action is taken immediately after the rth sample is $\mathbf{1}'(\mathbf{p}_{r-1} - \mathbf{p}_r)$, so that the average run length is given by

$$L = \mathbf{1}' \sum_{r=1}^{\infty} r(\mathbf{p}_{r-1} - \mathbf{p}_r) N,$$

i.e.
$$L = \mathbf{1}'(\mathbf{I} - \mathbf{P})^{-1} \mathbf{p}_0 N, \tag{7}$$

where N is the size of the sample.

In the general case the size of the matrix $\mathbf{P}$ increases rapidly with n so that the labour involved in inverting $\mathbf{I} - \mathbf{P}$ is considerable. Fortunately, the transition matrices for the rules III and IV are of a simple type and the states can be more simply specified so that the average run lengths can be found with little trouble. Let the probabilities that a given point falls between the warning lines, between the warning and action lines, and outside the action lines be p_0, p_1, p_2 respectively. Suppose that a suitable convention is made for points falling on the lines; for example, that any point falling on a line is to be regarded as falling into the adjacent more extreme region of the chart. With such a convention, clearly $p_0 + p_1 + p_2 = 1$.

Consider first rule III; it has been shown that this rule is equivalent to a sequence of linear sequential tests with initial score on the acceptance boundary. By inspection it is seen that appropriate scores to assign to the sample point are given by

$$a = 1, \quad b = n-1, \quad c = n, \tag{8}$$

and action is to be taken if the total penalty score rises a height $h = n$ above its previous least value; or equivalently if the sequential test ends on the rejection boundary given by

$h = n$. When action is not required the position of the total penalty score relative to its previous minimum or of the cumulative score in the test is specified by one of the integers $0, 1, \ldots, n-1$. These may be regarded as the states of a Markov chain and the transition matrix P is seen to be

$$P = \begin{bmatrix} p_0 & 0 & \cdots & \cdots & 0 & p_1 \\ p_0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & p_0 & 0 & \cdots & 0 & 0 \\ & & \cdots & & & \\ 0 & 0 & \cdots & \cdots & p_0 & 0 \end{bmatrix}. \tag{9}$$

The average run length is then given by equation (7). Alternatively, let L_i be the average number of observations drawn before action is taken when the state is i. Then L_0 is the average run length of rule III. We have, by taking expectations conditional upon the result of the first sample,

$$L_0 = p_0(N + L_0) + p_1(N + L_{n-1}) \tag{10}$$
$$= N + p_0 L_0 + p_1 L_{n-1}.$$

Similarly, we obtain the set of equations

$$\left. \begin{aligned} L_0 &= N + p_0 L_0 + p_1 L_{n-1} \\ L_1 &= N + p_0 L_0 \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ L_i &= N + p_0 L_{i-1} \end{aligned} \right\} \quad (1 \leqslant i < n-1). \tag{11}$$

The last $n-1$ of these equations may be solved successively for $L_1, \ldots, L_{n-1}$ in terms of L_0, and the first equation used to evaluate L_0. We obtain the average run length of rule III:

$$L = \frac{(1 - p_0 + p_1 - p_1 p_0^{n-1}) N}{(1 - p_0)(1 - p_0 - p_1 p_0^{n-1})} \quad (n \geqslant 2). \tag{12}$$

Scores for the various positions of the sample point on the chart for the repeated sequential tests equivalent to rule IV can be taken to be $a = n-1, b = 1, c = n$ and $h = n$. There are again only n possible states before action is taken for the Markov chain. The average run length is L_0, where L_0 is given by the equations

$$\left. \begin{aligned} L_i &= N + p_0 L_0 + p_1 L_{i+1} \\ L_{k-1} &= N + p_0 L_0 \end{aligned} \right\} \quad (0 \leqslant i < k-1). \tag{13}$$

We obtain the average run length of rule IV:

$$L = \frac{(1 - p_1^n) N}{1 - p_0 - p_1 + p_0 p_1^n}. \tag{14}$$

4. CHOICE OF SCHEME

In each of the two types of rule, III and IV, there are certain constants that may be chosen to give the scheme adopted some desired properties. Consider a two-sided chart with two sets of warning and action lines symmetrically placed about the ideal value for the parameter, or a one-sided chart with one warning and one action line. The disposable constants are then the sample size, N, the number of sample points considered, n, and the positions of the warning and action lines; the scheme may also be of one of two types. This may be contrasted with the single-sample control chart scheme in which there are only two disposable

constants, the sample size and the position of the action lines. In the single-sample case
the scheme could be chosen to give a specified average run length for some unsatisfactory
value of the parameter and maximum average run length for the ideal value (Page, 1954 b);
that is to say, the amount of scrap at a certain value of the parameter that can be tolerated
is stated and the scheme is chosen to satisfy this requirement and to give the longest run
without action being taken when the quality is satisfactory. A corresponding choice for
a scheme with warning lines could be made by selecting two unsatisfactory values of the
parameter and the average run lengths that can be tolerated at these values and then
choosing the disposable constants to gain maximum average run length on the ideal quality;
in this way the average run length will have desirable properties over a range of the para-
meter. In order to illustrate this we consider schemes for controlling the mean of a normal
population with known variance, σ^2, for two-sided deviations from the ideal mean. Suppose
that it is desired to obtain a scheme with average run lengths of 100 and 25 when there are
departures from the ideal mean of amounts 0.4σ and 0.8σ respectively. The large number of
disposable constants would make it laborious to find the scheme of the types considered
which has the greatest run length on ideal quality, but one scheme with approximately the
run lengths stated above is that of type IV with $k = 4$, sample size $N = 20$, warning lines
at $\mu \pm 1.5\sigma/\sqrt{N}$, and action lines at $\mu \pm 2.875\sigma/\sqrt{N}$. We can compare this scheme with the
best single-sample schemes for detecting the two sizes of departure from the mean. From
the tables given in an earlier paper (Page, 1954 b, Tables 2 a–c) the single-sample scheme
having average run length 100 when the mean is $\mu \pm 0.4\sigma$ and maximum average run length
when the mean is μ has action lines drawn at $\mu \pm 2.82\sigma/\sqrt{N}$, and sample size $N = 70$. Simi-
larly, the single-sample scheme with action lines at $\mu \pm 2.83\sigma/\sqrt{N}$ and sample size $N = 17$
has $L = 25$ at mean $\mu \pm 0.8\sigma$ and maximum average run length at mean μ. The average run
lengths for the three schemes for several values of λ, where the mean is $\mu \pm \lambda\sigma$, are shown in
Table 1. It is seen that the rule of type IV has approximately the same run length as the
single-sample scheme with sample size 17 for large deviations ($> 0.8\sigma$) from the true mean;
on the other hand, its run length for smaller deviations ($0.2 < |\lambda| < 0.4$) is rather less than
that of the single-sample scheme. Again it is seen that, while the type IV scheme compares
unfavourably with the second single-sample scheme ($N = 70$, $B_1 = 2.82$) in its behaviour

Table 1. *Average run lengths of control chart schemes*

λ	Single-sample schemes		Type IV scheme
	$N = 17$ $B_1 = 2.83$	$N = 70$ $B_1 = 2.82$	$n = k = 3$ $N = 20$ $B_1 = 2.875$ $B_2 = 1.5$
0·0	3652	14600	3368
0·2	780	560	545
0·4	146	100	99
0·6	47	71	42
0·8	25	70	26
1·0	19	70	21

on good quality, it gains considerably for the large departures ($|\lambda| > 0.4$); clearly for a single-sample scheme to cause action to be taken at least one sample point must be plotted. The situation we are considering is where an amount Nf of production is output between visits from the inspector (where f is the fraction of output sampled), thus the relevant quantity to consider for even extreme deviations from the mean is the average run length, or approximately the sample size.

The choice of the best sample size could be examined in the same way as in the earlier paper, but it would be extremely laborious to do so. In general, there will exist a value of N giving maximum average run length when the production has the ideal quality and specified average run lengths at two values of the mean. The effects of different sample sizes are illustrated by examples of rules of type IV which have average run lengths of 100 and 25 at $\mu \pm 0.4\sigma$, $\mu \pm 0.8\sigma$ respectively; the sample size, N, the positions of the warning and action lines, B_1 and B_2, the number, n, of consecutive samples considered, and the average run length, L, on the ideal mean are shown for five schemes in Table 2. There is, of course, an upper limit to the size of the sample that can be used for the scheme to have the desired run lengths. It is, however, again preferable to use larger samples if this is practically possible.

Table 2. *Average run lengths of type IV schemes*

N	n	B_1	B_2	$L(0.8\sigma)$	$L(0.4\sigma)$	$L(0)$
5	3	3.25	1.25	25.3	118.0	584.6
10	3	3.125	1.25	25.2	103.9	1096.9
15	3	3.00	1.375	25.6	104.2	2286.4
20	3	2.75	1.75	25.0	101.6	3155.9
23	3	2.625	2.0	25.9	92.3	2607.8

5. THE TABLES

In the Appendix are given tables of the average run lengths of schemes for controlling the mean of a normal population with known standard deviation. The run lengths are tabulated for some schemes of type IV with sample sizes 5, 10, 15 and 20, taking into account the means of the last three or four samples drawn ($n = 3$ or $n = 4$). In order to use the tables to determine the suitable control chart scheme to be used it is necessary to decide first upon the average run lengths that can be permitted at two process means different from the ideal one, and which of the above sample sizes it is most convenient to use. A scheme approximately satisfying these requirements may then be selected by inspection of the tables, possibly using some rough interpolation. The schemes shown in Tables 1 and 2 serve as examples of the method.

It has been remarked that the use of moderately sized samples improves the average run-length function; consequently such samples are to be preferred where possible. For samples of ten and over the average run length for departures from the ideal mean of more than a standard deviation is near the sample size; this is easily seen since the probability that a point falls outside the action lines is nearly unity.

Similar tables could be constructed for the average run lengths of schemes of type III. The two types of scheme are, of course, identical when the number of sample points

considered is two. A few calculations indicate that the two types of scheme have very similar properties and that, corresponding to a given scheme of one type, there exist schemes of the other type with approximately the same average run lengths.

6. SIMULTANEOUS CONTROL OF THE MEAN AND STANDARD DEVIATION

It is often required to control simultaneously both the mean and the standard deviation of a normal population. For example, when goods are being manufactured to a specification a change in either the process mean or standard deviation would alter the fraction of defective articles produced, one process inspection scheme for this situation (Jennett & Welch, 1939) is based upon the ratio $(U - x)/s$, where U is an upper tolerance limit, and x, s^2 sample estimates of mean and variance. Another possibility is to use the easily calculated estimate of the standard deviation from the sample range, w, in place of s and so base the scheme upon $(U - \bar{x})/w$. The more usual procedure in practice is to keep two charts, one recording the means and the other the ranges, of the samples. Before setting up any of these schemes it is important to consider both the type of rectifying action that will be taken and the changes that are likely to occur. If changes in the mean and standard deviation require different rectifying actions it is necessary that the inspection scheme indicate the type of change and not merely its existence. Again, the choice of scheme will be influenced by the relative frequency of occurrence and importance of the types of change. For example, if changes in the standard deviation are very infrequent there is little point in estimating σ from each sample as would be necessary in Jennett & Welch's rule. The same considerations will hold to decide the average run lengths for various types of change that the scheme must achieve. These remarks lead us to consider the possibility of controlling both mean and standard deviation on a single chart for means when changes in the standard deviation are relatively rare or unimportant. The remainder of this paper is devoted to the development of such a scheme and to comparing it with a conventional scheme.

The control chart scheme using charts for the mean and range are often of the conventional type with action lines drawn so that the probability that a sample point falls outside the action lines on a given chart when the parameters have their ideal values is an assigned amount, α ($\alpha \doteq 1/500$ gives the conventional 3-sigma limits on the mean chart). Alternatively and preferably, the action limits could be determined for each chart separately on the basis of the average run length as described above and elsewhere (Page, 1954b), so that the losses incurred by the scheme may be estimated. This approach, of course, will not be very accurate for some values of the parameters since, for example, a change in the standard deviation affects the run length of both charts. Consider a single-sample scheme for both charts so that action is taken if the mean of any sample, x, falls outside the range $(\mu - B\sigma/\sqrt{N},$ $\mu + B\sigma/\sqrt{N})$, or if the range w is greater than $r_1 \sigma$, where r_1 is suitably chosen. Since the mean and range of samples from a normal population are independently distributed with frequency elements $f(x \mid \mu', \sigma') d\bar{x}$, $g(w \mid \sigma') dw$, where μ', σ' are the process mean and standard deviation respectively, the probability that a given sample does not cause action to be taken is q_1', q_2', where

$$q_1' = \int_{\mu - B\sigma/\sqrt{N}}^{\mu + B\sigma/\sqrt{N}} f(\bar{x} \mid \mu', \sigma') d\bar{x} = 1 - p_1', \tag{15}$$

$$q_2' = \int_0^{v_1 \sigma} g(w \mid \sigma') dw = 1 - p_2. \tag{16}$$

Clearly q_1 q_2 are the probabilities that the mean and range respectively fall within the action lines. The run length of the charts together is distributed as the smaller of two independent geometric variates with parameters p_1, p_2 accordingly. The combined run length, l, is a geometric variable with parameter $1 - q_1' q_2'$.

$$P\{l = rN\} = (q_1' q_2')^{r-1} (1 - q_1' q_2').$$ (17)

The average run length is thus
$$L = \frac{N}{1 - q_1' q_2'}$$ (18)

or in terms of the average run length of the charts separately is

$$L = \frac{L_1 L_2}{L_1 + L_2 - 1},$$ (19)

with an obvious notation. A numerical example is given in the next section.

We turn now to investigate what control can be achieved by the use of the mean chart only, thus avoiding the need to calculate and plot ranges. A possible scheme makes use of the warning lines in another way. Consider the rule.

V. *Choose n, N. Plot the means of samples of size N on a chart on which are drawn two warning and two action lines. Take action if:*

(i) *any point falls outside the action lines,*

or (ii) *n consecutive points fall outside the warning lines,*

or (iii) *two out of any set of n consecutive points fall outside opposite warning lines.*

A sequence of n points outside a warning line is evidence that the process mean has moved in the corresponding direction: similarly, the occurrence of the means of two near samples outside opposite warning lines points to an increase in the spread of the distribution. This scheme differs from one based on the range in that it depends on the variation between samples and not on that within samples. Such a scheme cannot be expected to be very sensitive in detecting increases in the standard deviation and will become less sensitive as the sample size increases; however, it may serve to keep a check on the standard deviation when control of the mean is of prime importance. In order to evaluate the average run length, L_0, of rule V let $L_i'(L_i'')$ $i = 1, \ldots, n - 1$, be the average further number of articles drawn before action is taken when the last i sample points have fallen between the upper (lower) warning and action lines.

Let the probabilities that a sample point falls between the upper (lower) warning and action lines be $r(s)$. Then in the notation of § 3 we have $r + s = p_1$. By considering expectations conditional upon the result of the first sample we have

$$L_0 = p_0(L_0 + N) + r(L_1' + N) + s(L_1'' + N) + p_2 N.$$

Therefore
$$L_0 = p_0 L_0 + r L_1' + s L_1'' + N.$$ (20)

Again, by considering expectations conditional on the result of the next sample when the last i points have fallen between the upper warning and action lines, we obtain

$$L_i' = p_0 L_0 + r L_{i+1}' + N \quad (i = 1, 2, \ldots, n-2),$$ (21)

$$L_{n-1}' = p_0 L_0 + N.$$ (22)

A similar set of equations is obtained for the L_i''. Hence we have

$$L_1' = \frac{(N + p_0 L_0)(1 - r^{n-1})}{1 - r}, \tag{23}$$

and similarly

$$L_1'' = \frac{(N + p_0 L_0)(1 - s^{n-1})}{1 - s}. \tag{24}$$

These values substituted in (20) yield the solution

$$L_0 = \frac{(1 - rs - r^n - s^n + sr^n + rs^n) N}{p_2 + rs(1 + p_0) + p_0(r^n + s^n - sr^n - rs^n)}. \tag{25}$$

A similar rule for simultaneously controlling the mean and standard deviation is provided by rule III without modification. If the two 'warning' points causing action to be taken are on the same side of the ideal mean a shift in the mean in that direction will be suspected, while if they are on opposite sides an increase in the standard deviation will be suspected. Of course with this rule, if a change in the standard deviation occurs and action is taken because of points outside the warning lines, the two points are just as likely to be on the same side of the ideal mean as they are to be on opposite sides. Again, action could be taken after a change in the standard deviation because of a point outside the action lines. Consequently the wrong sort of change is more likely to be suggested by the scheme when a change in the standard deviation happens. This is of little importance if the rectifying action is independent of the suspected type of change. However, the correct inference is more likely to be drawn with rule V than with rule III.

The rules of this section have nothing corresponding to the lower control limit for the range sometimes drawn in the conventional method. The purpose of such a line is to enable the occurrences of samples with small range to be investigated in the hope that it will be possible to reduce the process standard deviation. It would, of course, be possible to introduce a condition calling for investigation if two successive samples both had means differing little from the required process mean: however, we shall not consider the consequences of such a complication.

7. NUMERICAL EXAMPLES OF SIMULTANEOUS CONTROL

Consider schemes to control the mean and standard deviation of a normal population. Suppose that both mean and range charts with action lines only are kept so that the average run length, L, of the scheme is given by equations (18) and (19). If the action lines are chosen so that the process inspection schemes formed by taking each of the charts on its own have equal average run lengths when both the process mean and standard deviation are at their ideal values, then the average run length of the rule using both charts with these parameter values is approximately half that of the separate rules. For different values of the mean while the standard deviation is the same, the run length of the range chart is unchanged; consequently the run length of the combination is a little less than that of the mean chart, and approaches it as the change in the mean increases. On the other hand, for different values of the standard deviation but the same mean, both L_1 and L_2 are changed. The average run lengths for a specific scheme with mean μ', and standard deviation σ' are shown in Tables 3 and 4. Let the action limits be drawn in the 'conventional' positions, i.e. so that the probability that a sample taken from ideal quality gives a point outside the action

line is 1 500 for each chart. Thus the lines on the mean chart are drawn at $\mu \pm 3.09\sigma / N$, and for samples of $N = 5$ the range chart has a line at $w_1 = 5.24\sigma$. We then have

$$q_1 = \int_{(-B-\lambda\sqrt{N})/K}^{(B-\lambda\sqrt{N})/K} \frac{1}{\sqrt{(2\pi)}} \cdot e^{-\frac{1}{2}t^2} dt, \tag{26}$$

$$q_2 = \int_0^{w_1/K\sigma} g(w)\, dw. \tag{27}$$

For comparison, the average run lengths of one of the combined rules are shown in the tables, the scheme chosen has the same sample size, $N = 5$, and $B_1 = 3.00$, $B_2 = 2.00$, $n = 3$. With these values the combined rule has approximately the same average run length on the ideal quality as the scheme based upon both mean and range charts. It is seen that the rules gives fair control against changes in the standard deviation, and for changes in the mean has a smaller average run length than the rule using the two charts. Consequently when it is necessary only to keep an eye on the standard deviation while controlling the mean of a normal distribution a scheme of type V may be suitable. If so, the labour of keeping both mean and range charts can be avoided.

Table 3. *Average run lengths of schemes.* $N = 5$, $\mu' = \mu + \lambda\sigma$, $\sigma' = \sigma$

$|\lambda|$	Rule using			Combined rule
	Mean	Range	Mean and range	
0·0	2500	2500	1250	1383
0·2	1149	2500	789	767
0·4	359	2500	314	257
0·6	125	2500	119	90
0·8	52	2500	51	38
1·0	25	2500	25	20

Table 4. *Average run lengths of schemes,* $N = 5$, $\mu' = \mu$, $\sigma' = K\sigma$

K	Rule using		Combined rule
	Range	Mean and range	
1·0	2500	1250	1383
1·25	198	158	241
1·5	51	38	91
1·75	24	18	50
2·0	15	12	33
2·25	11	9	25
2·5	9	7	20

A short table of schemes of type V is given in the Appendix, Table 2, for controlling a normal population with samples of five measurements, using at most a sequence of three such samples. The average run lengths for changes in the mean are shown in Table 2*a*, and for changes in the standard deviation in Table 2*b*.

I wish to thank Dr D. R. Cox for many helpful discussions on the subject of this paper, and the Director, Mathematical Laboratory, Cambridge, for permission to use the EDSAC for the calculation of the tables.

REFERENCES

Aroian, L. A. & Levene, H. (1950). *J. Amer. Statist. Ass.* **45**, 520.

Barnard, G. A. (1946). *J.R. Statist. Soc. Suppl.* **8**, 1.

Bartlett, M. S. (1953). *Proc. Camb. Phil. Soc.* **49**, 263.

Dudding, B. P. & Jennett, W. J. (1944). *Quality Control Chart Technique.* London: General Electric.

Feller, W. (1950). *Probability Theory and its Application.* New York: Wiley.

Jennett, W. J. & Welch, B. L. (1939). *J.R. Statist. Soc. Suppl.* **6**, 80.

Page, E. S. (1954*a*). *Biometrika*, **41**, 100.

Page, E. S. (1954*b*). *J.R. Statist. Soc.* B, **16**, 131.

Shewhart, W. A. (1931). *Economic Control of Quality of Manufactured Product.* New York: Macmillan.

Wald, A. (1947). *Sequential Analysis.* New York: Wiley.

Weiler, H. (1953). *J. Amer. Statist. Ass.* **48**, 816.

Tables of average run lengths of rules IV

Charts with action lines at $\mu \pm B_1\sigma\sqrt{N}$, warning lines at $\mu \pm B_2\sigma\sqrt{N}$.

Table 1a. $N = 5$

	B_2	$n=3$					$n=4$					B_1
λ		1·00	1·25	1·50	1·75	2·00	1·00	1·25	1·50	1·75	2·00	λ
$B_1 =$ 3·00	0·0	202	368	1074	1398	1802	727	1490	1692	1920	1850	0·0
	0·2	137	285	527	751	875	96	98	95	552	887	0·2
	0·4	63	102	160	221	263	42	135	256	268	240	0·4
	0·6	32	43	57	75	90	41	63	29	95	100	0·6
	0·8	19	23	27	32	38	26	29	34	39	43	0·8
	1·0	14	15	16	18	20	16	17	19	20	21	1·0
	1·2	11	11	12	13	13	12	12	12	13	13	1·2
	1·4	9	9	9	9	9	8	8	8	9	9	1·4
	1·6	7	7	7	7	7	7	7	7	7	7	1·6
	1·8	6	6	6	6	6	6	6	6	6	6	1·8
$B_1 =$ 3·125	0·0	208	549	1326	2251	2691	579	1913	2451	2758	2805	0·0
	0·2	142	312	637	1008	1213	337	723	1100	1254	1253	0·2
	0·4	66	111	184	274	311	122	206	300	371	382	0·4
	0·6	33	45	63	87	110	51	70	94	115	125	0·6
	0·8	20	24	29	36	44	27	32	39	45	51	0·8
	1·0	14	16	17	19	22	18	19	21	23	25	1·0
	1·2	11	12	12	13	13	13	13	13	14	14	1·2
	1·4	9	9	9	9	9	9	9	9	9	9	1·4
	1·6	7	7	7	7	7	7	7	7	7	7	1·6
	1·8	6	6	6	6	6	6	6	6	6	6	1·8
$B_1 =$ 3·25	0·0	213	585	1575	3110	4041	621	1854	3518	4202	4319	0·0
	0·2	146	334	747	1331	1726	364	873	1594	1810	1881	0·2
	0·4	68	118	207	335	452	132	237	378	183	527	0·4
	0·6	35	48	69	100	134	54	75	110	113	165	0·6
	0·8	21	25	31	40	50	29	35	43	53	62	0·8
	1·0	15	16	18	21	24	19	21	23	25	28	1·0
	1·2	12	12	13	13	14	14	14	15	15	16	1·2
	1·4	9	9	10	10	10	10	10	10	10	11	1·4
	1·6	8	8	8	8	8	8	8	8	8	8	1·6
	1·8	6	6	6	6	6	6	6	6	6	6	1·8

Table 1b. $N = 10$

		$n=3$					$n=4$					B_2
	B_2	1·00	1·25	1·50	1·75	2·00	1·00	1·25	1·50	1·75	2·00	λ
	λ											
$B_1 =$ 2·875	0·0	387	896	1684	2249	2434	933	816	2333	2456	2473	0·0
	0·2	191	336	531	689	763	360	572	720	774	786	0·2
	0·4	70	90	119	149	171	98	127	156	175	183	0·4
	0·6	34	38	43	49	55	42	46	51	56	59	0·6
	0·8	22	22	23	24	26	24	25	25	26	27	0·8
	1·0	15	16	16	16	16	16	16	16	16	16	1·0
$B_1 =$ 3·00	0·0	404	1005	2149	3197	3604	1055	2380	3384	3659	3700	0·0
	0·2	201	374	643	904	1044	405	709	963	1069	1095	0·2
	0·4	71	97	134	177	213	107	146	188	220	236	0·4
	0·6	35	40	47	55	63	45	51	58	65	70	0·6
	0·8	23	24	25	26	28	26	27	28	29	30	0·8
	1·0	16	16	17	17	17	17	17	17	17	18	1·0
$B_1 =$ 3·125	0·0	417	1097	2651	4502	5382	1159	3026	4901	5516	5613	0·0
	0·2	209	408	761	1174	1436	447	862	1285	1491	1545	0·2
	0·4	74	104	149	208	264	116	165	225	277	306	0·4
	0·6	37	43	51	61	72	48	56	65	75	84	0·6
	0·8	24	25	27	29	31	28	29	31	32	34	0·8
	1·0	17	17	18	18	18	18	19	19	19	19	1·0
$B_1 =$ 3·25	0·0	426	1169	3150	6220	8081	1241	3707	7036	8404	8637	0·0
	0·2	216	437	879	1502	1980	483	1022	1700	2094	2208	0·2
	0·4	76	110	164	241	325	124	184	266	348	399	0·4
	0·6	39	45	54	67	82	51	60	73	87	100	0·6
	0·8	25	27	28	31	34	30	32	33	36	38	0·8
	1·0	18	19	19	19	20	20	20	20	20	21	1·0

APPENDIX (cont.)

Table 1c. $N = 15$

B_1	λ	$n=3$					$n=4$					B_2
$B_2 \rightarrow$		1·00	1·25	1·50	1·75	2·00	1·00	1·25	1·50	1·75	2·00	λ
$B_1=$ 2·75	0·0	548	1163	1929	2367	2492	1201	2034	2423	2505	2516	0·0
	0·2	208	324	457	556	601	344	483	573	607	615	0·2
	0·4	67	80	95	110	123	86	100	114	124	129	0·4
	0·6	35	36	38	40	42	39	40	42	43	44	0·6
	0·8	23	23	23	23	23	23	23	23	24	24	0·8
	1·0	17	17	17	17	17	17	17	17	17	17	1·0
$B_1=$ 2·875	0·0	581	1344	2526	3373	3651	1399	2724	3500	3684	3710	0·0
	0·2	222	365	552	714	800	392	594	748	811	831	0·2
	0·4	71	86	106	128	147	94	114	134	149	158	0·4
	0·6	37	39	41	44	47	42	44	46	48	50	0·6
	0·8	24	24	24	24	25	25	25	25	25	25	0·8
	1·0	18	18	18	18	18	18	18	18	18	18	1·0
$B_1=$ 3·00	0·0	607	1508	3223	4795	5406	1582	3570	5076	5489	5549	0·0
	0·2	234	403	656	911	1071	439	722	977	1101	1137	0·2
	0·4	74	93	117	147	175	103	127	156	181	196	0·4
	0·6	39	41	44	48	53	46	48	51	54	57	0·6
	0·8	25	25	26	26	27	27	27	27	27	27	0·8
	1·0	18	18	18	18	18	18	18	18	18	18	1·0
$B_1=$ 3·125	0·0	626	1645	3977	6753	8073	1738	4539	7351	8275	8420	0·0
	0·2	244	438	765	1159	1439	482	862	1270	1502	1574	0·2
	0·4	78	98	128	167	208	110	141	180	218	244	0·4
	0·6	41	44	48	53	58	49	52	56	61	65	0·6
	0·8	27	27	27	28	29	29	29	29	30	30	0·8
	1·0	19	19	19	19	19	19	19	19	19	19	1·0

Table 1d. $N = 20$

B_1	λ	$n=3$					$n=4$					B_2
$B_2 \rightarrow$		1·00	1·25	1·50	1·75	2·00	1·00	1·25	1·50	1·75	2·00	λ
$B_1=$ 2·75	0·0	730	1550	2571	3156	3323	1601	2712	3230	3340	3355	0·0
	0·2	226	333	458	558	608	355	484	575	614	625	0·2
	0·4	70	79	90	102	111	86	95	105	112	117	0·4
	0·6	37	38	39	40	41	40	40	41	41	42	0·6
	0·8	25	25	25	25	25	25	25	25	25	25	0·8
	1·0	21	21	21	21	21	21	21	21	21	21	1·0
$B_1=$ 2·875	0·0	774	1792	3368	4497	4868	1896	3632	4667	4913	4947	0·0
	0·2	211	372	545	703	795	402	586	737	810	832	0·2
	0·4	71	85	99	115	129	93	106	120	132	140	0·4
	0·6	39	40	41	43	45	43	44	45	46	46	0·6
	0·8	26	26	26	26	26	26	26	26	26	26	0·8
	1·0	21	21	21	21	21	21	21	21	21	21	1·0
$B_1=$ 3·00	0·0	809	2010	4297	6394	7208	2110	4761	6768	7318	7399	0·0
	0·2	254	469	678	882	1045	447	701	943	1077	1121	0·2
	0·4	78	91	108	129	150	101	117	136	155	168	0·4
	0·6	42	43	45	47	49	47	48	49	50	52	0·6
	0·8	25	27	27	27	28	28	28	28	28	28	0·8
	1·0	22	22	22	22	22	22	22	22	22	22	1·0
$B_1=$ 3·125	0·0	831	2194	5362	9003	10764	2317	6053	9802	11033	11226	0·0
	0·2	264	443	735	1095	1376	489	824	1200	1440	1528	0·2
	0·4	81	96	117	144	174	109	129	155	182	203	0·4
	0·6	44	46	48	50	53	50	52	53	55	57	0·6
	0·8	29	29	29	29	29	29	29	29	30	30	0·8
	1·0	22	22	22	22	22	22	22	22	22	22	1·0

APPENDIX (cont.)

Tables of average run lengths of rules V

Charts with action lines at $\mu \pm B_1\sigma/\sqrt{N}$, warning lines at $\mu \pm B_2\sigma/\sqrt{N}$. Sample size, $N = 5$.

Table 2a. $\mu' = \mu + \lambda\sigma$, $\sigma' = \sigma$.

	B_2	1·50	1·75	2·00
	λ			
$B_1 = 2\cdot875$	0·0	404	722	1021
	0·2	280	457	564
	0·4	128	170	196
	0·6	51	63	73
	0·8	25	29	32
	1·0	15	16	17
$B_1 = 3\cdot00$	0·0	446	879	1383
	0·2	344	577	767
	0·4	148	211	257
	0·6	57	74	90
	0·8	27	32	38
	1·0	16	18	20
$B_1 = 3\cdot125$	0·0	481	1033	1830
	0·2	385	713	1037
	0·4	169	260	336
	0·6	63	86	110
	0·8	29	36	44
	1·0	17	19	22
$B_1 = 3\cdot25$	0·0	507	1173	2340
	0·2	422	856	1384
	0·4	188	314	438
	0·6	69	99	133
	0·8	32	40	50
	1·0	18	21	24

Table 2b. $\mu' = \mu$, $\sigma' = K\sigma$.

	B_1	1·50	1·75	2·00
	K			
$B_1 = 2\cdot875$	1·00	404	722	1021
	1·25	117	160	198
	1·50	56	69	79
	1·75	35	41	45
	2·00	26	29	30
	2·25	20	22	23
$B_1 = 3\cdot00$	1·00	446	879	1383
	1·25	129	186	241
	1·50	61	77	91
	1·75	38	44	50
	2·00	27	31	33
	2·25	22	23	25
$B_1 = 3\cdot125$	1·00	481	1033	1830
	1·25	140	213	292
	1·50	66	86	105
	1·75	41	48	56
	2·00	29	33	36
	2·25	23	25	27
$B_1 = 3\cdot25$	1·00	507	1173	2340
	1·25	151	210	349
	1·50	71	95	120
	1·75	44	53	62
	2·00	31	35	39
	2·25	24	26	29

MISCELLANEA

Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's t-distribution[*]

By C. W. DUNNETT[†] and M. SOBEL[‡]

Cornell University

In a recent paper (Dunnett & Sobel, 1954), a multivariate analogue of Student's t-distribution was defined as the joint distribution of p variates $t_{in} = z_i/s$ $(i = 1, 2, ..., p)$. Here the z_i have a non-singular multivariate normal distribution with means 0, common unknown variance σ^2 and known correlation matrix $\{\rho_{ij}\}$ and ns^2/σ^2 has a χ^2-distribution, independent of the z_i, with n degrees of freedom. The joint density of the t_{in} is given by

$$f(t_{1n}, ..., t_{pn}) = \frac{A^{\frac{1}{2}}\Gamma\{\frac{1}{2}(n+p)\}}{(n\pi)^{\frac{1}{2}p}\Gamma(\frac{1}{2}n)} \left[1 + \frac{1}{n}\sum_{i,j} a_{ij} t_{in} t_{jn}\right]^{-\frac{1}{2}(n+p)}, \tag{1}$$

where A is the determinant of the positive definite matrix $\{a_{ij}\} = \{\rho_{ij}\}^{-1}$. In the authors' previous paper expressions and tables for the probability integral and equi-co-ordinate percentage points of (1) were obtained for the bivariate case ($p = 2$). In that paper an equi-co-ordinate P-percentage point was defined as the value of h for which

$$\int_{-\infty}^{h} ... \int_{-\infty}^{h} f(t_{1n}, ..., t_{pn}) dt_{1n} ... dt_{pn} = P. \tag{2}$$

In this note we shall derive approximations (which are also lower bounds) to the probability integral of (1) applicable in special cases when $p > 2$. These results then can be used to obtain approximations (which are also upper bounds) to any equi-co-ordinate P-percentage point. Equation (11) below shows that the probability integral table (Table 1) of the previous paper can be used when $\rho_{ij} = \frac{1}{2}$ $(i \neq j)$ to obtain numerically the approximations referred to above.

Letting $I = I_n(h_1, h_2, ..., h_p)$ denote the left-hand member of (2) with the upper limit of t_{in} replaced by h_i $(i = 1, 2, ..., p)$, we have

$$I = \Pr\{t_{1n} < h_1, ..., t_{pn} < h_p\} = \Pr\{z_1 < h_1 s, ..., z_p < h_p s; n\}. \tag{3}$$

Fixing s as the last variable to be integrated, we can write

$$I = \int_0^{\infty} G_p(h_1 s, ..., h_p s; \{\rho_{ij}\}) f_n(s) ds, \tag{4}$$

where $G_p = G_p(x_1, ..., x_p; \{\rho_{ij}\})$ is the c.d.f. of the standardized p-variate normal distribution with correlation matrix $\{\rho_{ij}\}$, and $f_n(s)$ is the probability density function of s with n degrees of freedom.

ASSUMPTION 1. The matrix $\{\rho_{ij}\}$ has the structure $\rho_{ij} = b_i b_j$ $(i \neq j)$, where $0 \leq b_i < 1$ $(i = 1, 2, ..., p)$. It follows that $\{\rho_{ij}\}$ is positive definite, since the associated quadratic form $\sum_i (1 - b_i^2) x_i^2 + (\sum_i b_i x_i)^2$ is positive for x_i not all zero.

ASSUMPTION 2. The upper limits of integration in (3) are non-negative, i.e. $h_i \geq 0$ $(i = 1, 2, ..., p)$.

We note that when $\rho_{ij} = \rho \geq 0$ $(i \neq j)$, Assumption 1 is satisfied since this occurs when $b_i = \sqrt{\rho}$ $(i = 1, 2, ..., p)$.

If we let $y_0, y_1, ..., y_p$ denote independent, normally distributed chance variables with zero means and unit variances, and let $c_i = \sqrt{(1 - b_i^2)}$, then the joint distribution of the chance variables

$$z_i = c_i y_i - b_i y_0 \quad (i = 1, 2, ..., p)$$

is a standardized p-variate normal distribution with correlation matrix $\{\rho_{ij}\}$.

† Now at Lederle Laboratories, Pearl River, N.Y.

‡ Now at Bell Telephone Laboratories, Allentown, Pa.

[faded text] function G_p, and also [faded] and [faded] $i = 1, 2, \ldots$ Then

$$G_p = \Pr\{c_i y_i - b_i y_0 < a_i \quad (i = 1, 2, \ldots, p)\}$$

[faded display equation]

where [faded] is the standard univariate normal [faded]
[faded] it can easily be shown that, for [faded] Kimball 1951 that F_1 and F_2 are decreasing, bounded functions of I, [faded] $= 1, 2, \ldots$, of a chance variable [faded] letting E denote expectation,

$$E\left\{\prod_{i=1}^{r} F_i(x)\right\} \geq \prod_{i=1}^{r} E(F_i(x)). \tag{6}$$

Applying (6) to (5) gives

$$G_p \geq \prod_{i=1}^{p} \int_{-\infty}^{\infty} G\left(\frac{\cdots}{\cdots}\right) \, dy_0$$

$$= \prod_{i=1}^{p} \Pr\{c_i y_i - b_i y_0 < a_i\} = \prod_{i=1}^{p} G(a_i). \tag{7}$$

Substituting this result in (4) and applying (6) again gives

$$I \geq \int_0^{\infty} \prod_{i=1}^{p} G(h_i, s) f_n(s) \, ds \geq \prod_{i=1}^{p} \int_0^{\infty} G(h_i, s) f_n(s) \, ds$$

$$= \prod_{i=1}^{p} \Pr\{z_i < h_i s; n\} = \prod_{i=1}^{p} \Pr\{t_{in} < h_i\}. \tag{8}$$

This lower bound to I does not depend on $\{\rho_{ij}\}$ and is easily calculated from tables of the c.d.f. of the univariate Student t-distribution with n degrees of freedom.

Under Assumptions 1 and 2 a sharper (i.e. higher) lower bound can be obtained by obvious modifications of the above argument. This lower bound depends on the c.d.f. of the bivariate t-distribution considered by Dunnett & Sobel (1954) which is tabulated there for the special case $\rho = \frac{1}{2}$ and $h_1 = h_2 \geq 0$. The results are for even $p \geq 2$,

$$I \geq \prod_{i=1}^{\frac{1}{2}p} \Pr\{t_{2i-1, n} < h_{2i-1}, t_{2i, n} < h_{2i}\}, \tag{9}$$

and, for odd $p \geq 3$,

$$I \geq \Pr\{t_{1n} < h_1\} \prod_{i=1}^{\frac{1}{2}(p-1)} \Pr\{t_{2i, n} < h_{2i}, t_{2i+1, n} < h_{2i+1}\}. \tag{10}$$

If we replace Assumptions 1 and 2 above by

ASSUMPTION 1′. The matrix $\{\rho_{ij}\}$ has the structure $\rho_{ij} = \rho$ $(i \neq j)$, where $0 \leq \rho < 1$. Clearly, $\{\rho_{ij}\}$ is positive definite.

ASSUMPTION 2′. The upper limits of integration in (3) are all equal, i.e. $h_i = h$ $(i = 1, 2, \ldots, p)$;

then we can replace (9) and (10) by a single inequality and write for any integer $p \geq 2$

$$I \geq [\Pr\{t_{1n} < h, t_{2n} < h\}]^{\frac{1}{2}p}, \tag{11}$$

which is sharper than (10) (when p is odd). The proof of (11) is similar to the above proof, using, instead of (6), the well-known inequality (see, for example, Cramér, 1946, p. 176)

$$\beta_p^{1/p} \geq \beta_q^{1/q} \quad \text{for} \quad p \geq q, \tag{12}$$

where β_p denotes the pth absolute moment.

Paulson (1952) suggested an alternative method of obtaining an approximation and lower bound to (3) which is based on the Bonferroni inequality

$$\Pr\{t_{1n} < h_1, \ldots, t_{pn} < h_p\} \geq 1 - \sum_{i=1}^{p} \Pr\{t_{in} > h_i\}. \tag{13}$$

This method has the advantage that it requires no assumptions. However, when the inequalities (8) through (11) hold, then (8) and hence also (9), (10) and (11) give sharper lower bounds than (13). To show this for (8) let

$$1 - q_i = \Pr\{t_{in} < h_i\}, \tag{14}$$

so that $0 \leq q_i \leq 1$. Then we have to prove that

$$\prod_{i=1}^{p} (1 - q_i) \geq 1 - \sum_{i=1}^{p} q_i. \tag{15}$$

This certainly holds for $p = 1$, and a straightforward mathematical induction shows that (15) holds for all positive integers p.

The approximations (13), (8), (9), (10) and (11) can be used to obtain upper bounds to the equi-co-ordinate percentage points of (1) for $p > 2$. Table 1 compares the approximations with the exact values for the special cases $\rho_{ij} = \frac{1}{2}\,(i \neq j)$; $n = 5$, ∞; $p = 3$, 9 and $P = 0.50$, 0.75, 0.95, 0.99; all entries are rounded to the nearest two decimal places. For $n = \infty$ columns (13) and (8) require a table of the univariate normal c.d.f.; columns (10), (11) and the exact values were obtained from unpublished tables of the National Bureau of Standards (1953). For $n = 5$, columns (13) and (8) require a table of the univariate Student c.d.f. with 5 degrees of freedom; columns (10) and (11) and the exact values were computed by numerical integration, i.e. by applying Simpson's rule and using the National Bureau of Standards tables (1953) to evaluate the integral in (4). The exact values for $P = 0.95$, 0.99 will also appear in another paper by Dunnett where equi-co-ordinate percentage points for the case $\rho_{ij} = \frac{1}{2}\,(i \neq j)$ will be given for $p = 3(1)9$, $P = 0.95$, 0.99 and $n = 5(1)20$, 24, 30, 40, 60, 120, ∞.

Table 1. *Comparison of exact equi-co-ordinate percentage points with approximations for selected values of n, p and P*

P	n	$p=3$ Approximations				Exact values	$p=9$ Approximations				Exact values
		(13)	(8)	(10)	(11)		(13)	(8)	(10)	(11)	
0·99	5	4·46	4·46	4·37	4·32	4·21	5·75	5·75	5·61	5·59	5·03
	∞	2·71	2·71	2·70	2·70	2·68	3·06	3·06	3·05	3·05	3·00
0·95	5	2·91	2·90	2·82	2·78	2·68	3·93	3·90	3·79	3·78	3·30
	∞	2·13	2·12	2·10	2·09	2·06	2·54	2·53	2·52	2·51	2·42
0·75	5	1·62	1·55	1·46	1·42	1·32	2·48	2·38	2·28	2·26	1·81
	∞	1·38	1·33	1·28	1·26	1·19	1·91	1·86	1·82	1·82	1·60
0·50	5	1·07	0·89	0·80	0·74	0·62	1·93	1·71	1·60	1·58	1·10
	∞	0·97	0·82	0·74	0·70	0·59	1·59	1·45	1·39	1·38	1·04

For each of the approximations (13), (8), (10) and (11) in Table 1 it is conjectured that further calculation will establish the following properties for the difference $D = D(n, p, P)$ between the approximation and its exact value:

(i) D is increasing with p for each n and P,
(ii) D is decreasing with n for each p and P,
(iii) D is decreasing with P for $n = \infty$ and each p,
(iv) D is parabolic-shaped with P for $n = 5$ and each p.

The values of n, p and P in Table 1 were selected to cover a wide range of practical interest. Since only a limited number of exact values for finite n are known the inequalities considered in this paper, which have a fairly wide application, should prove to be useful.

REFERENCES

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.
DUNNETT, C. W. A multiple comparisons test for comparing several treatments with a control. Unpublished manuscript.
DUNNETT, C. W. & SOBEL, M. (1954). *Biometrika*, **41**, 153.
KIMBALL, A. W. (1951). *Ann. Math. Statist.* **22**, 600.
NATIONAL BUREAU OF STANDARDS (1953). Personal communication on unpublished tables.
PAULSON, E. (1952). *Ann. Math. Statist.* **23**, 239.

Galton's rank-order test

By J. L. HODGES, Jr.

University of California, Berkeley

1. One of the first uses of rank-order in statistics was that of Galton in the study of data referred to him by Charles Darwin (1876). A quantity was measured on each of n treated subjects and also on each of n control subjects, obtaining measurements $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ respectively. The $2n$ measurements were arranged in common increasing order of size, and Galton counted the number, say G, of times that an x of given rank exceeded the y of the same rank. In Galton's case, n was 15 and G was 13, which was regarded as evidence that the first sample came from a population stochastically larger than that from which the second sample came. In modern language, if G is sufficiently large, we reject the null hypothesis that the treatment is without effect, in favour of the alternative that the treatment tends to increase the measurements.*

2. Galton was not able to attach a significance level to his observation, inasmuch as he did not know the distribution of G under the null hypothesis. However, that distribution has recently been discovered in another connexion and proved to be very simple and elegant. Consider the usual penny-tossing game played by Peter and Paul, in which Peter pays one unit to Paul if the penny lands 'heads' and Paul pays one unit to Peter if it lands 'tails'. Suppose we are given that after $2n$ tosses, the contestants are even. Let F denote the number of times that Peter was in the lead (where conventionally we assert that, when the game is even, the player leads who led on the preceding toss). If we identify Peter's winning the kth toss with the event that the kth measurement in the Galton problem is an x, it appears that $F = 2G$.

The conditional distribution of F has been found by Chung & Feller (1949) to be uniform, under the hypothesis that Peter's n victories are randomly distributed among the $2n$ trials. But this is just the distribution of the ranks of the x measurements under the hypothesis that the treatment has no effect. Therefore we may assert, under the null hypothesis, that $P(G \leqq g) = (n - g + 1)/n$. For example, to Galton's observation we may attach the significance probability 1.5.

3. The proof of Chung & Feller is by means of a double generating function, and it may be of interest to have an enumerative proof of so simple a result. Consider the class of $\binom{2n}{n}$ possible arrangements of n 0's and n 1's, and for each such arrangement determine its score g by counting those 1's which precede the 0 of the same ordinal number. For example, in the sequence 01110001 the black 1's are counted, since the second 1 precedes the second 0, the third 1 precedes the third 0, but the first and fourth 1's are not counted as they follow the first and fourth 0's respectively. The sequence just given has a total score $g = 2$. As a sequence of length $2n$, the possible values of g are $0, 1, \ldots, n$. We shall denote the score of a sequence $(a_1 \ldots a_{2n})$ by $[a_1 \ldots a_{2n}]$.

THEOREM. *The number of arrangements with score x is independent of x.*

We shall prove this by defining a mapping T_n which has the following properties:

(a) The domain of T_n is the set of all sequences of length $2n$ whose score is positive.
(b) The range of T_n is the set of all sequences of length $2n$ whose score is less than n.
(c) T_n is a 1-1 function.
(d) $[T_n(a_1 \ldots a_{2n})] = [a_1 \ldots a_{2n}] - 1$.

As such a function gives a 1-1 mapping of the arrangements with score x on to those with score $x - 1$ for $x = 1, \ldots, n$, it establishes the desired result.

We shall need to consider the first point in the sequence at which there have been equal numbers of 0's and 1's (this corresponds in coin-tossing to the first toss at which the game is even). For each arrangement $(a_1, \ldots, a_{2n})$ let k be the smallest positive integer such that $a_1 + \ldots + a_{2k} = k$. We say that the sequence 'breaks' at $2k$, and note that

$$[a_1 \ldots a_{2n}] = [a_1 \ldots a_{2k}] + [a_{2k+1} \ldots a_{2n}]. \tag{1}$$

---

* Galton's analysis has been extensively reviewed by Fisher (1945, Ch. III), who points out that Darwin did not actually have two samples of n, but n matched pairs. It may be noted that the pairing did not serve to reduce the variance; in fact, if we test the hypothesis that there is no pair effect we have $F_{14, 14} = 0.554$. In any case, our present concern is with the problem of two samples of n.

Note that k may equal n. As $2k$ is the *first* equilibrium point, there cannot have been an earlier change of lead, so $[a_1 \ldots a_{2k}]$ must be either 0 or k. Further, if $[a_1 \ldots a_{2k}] = k$, we must have $a_1 = 1$, $a_{2k} = 0$ and $[a_2 \ldots a_{2k-1}] = k-1$.

We now define the functions T_n inductively. Let $T_1(10) = (01)$ and check properties (a)–(d) for this case. Suppose we have defined T_m for $m < n$ satisfactorily. For any $(a_1 \ldots a_{2n})$ of positive score, let $T_n(a_1 \ldots a_{2n})$ be

    (i) $(a_1 \ldots a_{2k} T_{n-k}(a_{2k+1} \ldots a_{2n}))$ if $[a_{2k+1} \ldots a_{2n}] > 0$,

    (ii) $(0 a_{2k+1} \ldots a_{2n} | a_2 \ldots a_{2k-1})$ if $[a_{2k+1} \ldots a_{2n}] = 0$.

We must check (a)–(d), of which (a) is obvious. Condition (d) holds for (i) by the induction hypothesis and (1). As for (ii), if $[a_{2k+1} \ldots a_{2n}] = 0$, we must have

$$0 < [a_1 \ldots a_{2k}] = k = [a_1 \ldots a_{2n}], \quad [0 a_{2k+1} \ldots a_{2n}] = 0, \quad [a_2 \ldots a_{2k-1}] = k - 1 = [T_n(a_1 \ldots a_{2n})].$$

From (d) it appears that the range of T_n is contained in the set of sequences with score less than n; any such sequence has a T_n-inverse under (i) if $[a_{2k+1} \ldots a_{2n}] < n - k$, since T_{n-k} satisfies the conditions, and under (ii) otherwise, since when $[a_{2k+1} \ldots a_{2n}] = n - k$ we have

$$k > [a_1 \ldots a_{2k}] = 0, \quad a_1 = 0, \quad a_{2k} = 1, \quad [a_2 \ldots a_{2k-1}] = 0, \quad (a_1 \ldots a_{2n}) = T_n(1 a_{2k+1} \ldots a_{2n} 0 a_2 \ldots a_{2k-1}).$$

The ranges of (i) and (ii) are disjoint and each is invertible, so (c) holds.

4. While the Galton test as defined above applies only when the two samples are of equal size, there is a natural extension to the more general case. For example, let $n_1 = 3$, $n_2 = 11$. The third, sixth, and ninth ranking observations in the larger sample divide it into four segments, each containing two observations. If we regard these three observations as 'representing' the larger sample, we may calculate G as before between the two samples of three each, and it is clear that G thus defined has the uniform distribution, since to each arrangement for it there corresponds the same number (27) of arrangements of the original problem. In general, if $n_2 = n_1 + k(n_1 + 1)$, we may represent the second sample by its observations of rank $k + 1, 2k + 2, \ldots, n_1 k + n_1$, and obtain a G which is uniformly distributed over the values $0, 1, \ldots, n_1$.

If $n_2 - n_1$ does not happen to be divisible by $n_1 + 1$, the above method will not apply exactly, but we may still obtain a uniformly distributed test statistic by randomization. The representative values of the larger sample may be chosen so that the segments into which they partition it differ by at most one in size. If we select one among all such partitionings at random, the resulting G may be shown to be uniformly distributed. Because of the natural repugnance of randomized decisions in such problems, it is probably preferable instead to associate with each arrangement the distribution of G values it has among the partitions, and choose for its definitive G the integer nearest the expectation of this distribution.

An interesting result follows if we let $n_2 \to \infty$. The population from which the second sample was chosen in then known, and we are dealing with the one sample rather than the two sample problem. Our statistic is now definable as the number of the quantities $y_k - F^{-1}(n_1 + 1 - k | n_1 + 1)$ which are positive. That its distribution is still uniform may be seen by a limiting argument from the above, or by considering it as a set of conditioned partial sums.

REFERENCES

CHUNG, KAI LAI & FELLER, W. (1949). On fluctuations in coin-tossing. *Proc. Nat. Acad. Sci., Wash.*, **35**, 605–8.

DARWIN, CHARLES (1876). *The Effects of Cross- and Self-fertilization in the Vegetable Kingdom*. London: John Murray.

FISHER, R. A. (1945). *The Design of Experiments*, 4th ed. Edinburgh: Oliver and Boyd.

On bounds for the normal integral[*]

By JOHN T. CHU

Institute of Statistics, University of North Carolina

1. Let
$$v = \int_0^x (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}t^2} dt \quad (x \geq 0). \tag{1}$$

G. Pólya (1949) and J. D. Williams (1946) proved independently that

$$v \leq \frac{1}{2}(1 - e^{-2x^2/\pi})^{\frac{1}{2}}. \tag{2}$$

Two simple questions follow naturally: (i) Is it possible to replace the constant $2/\pi$ or 2 by a smaller quantity without breaking the inequality? (ii) Does there exist a lower bound in a similar form for v? We find the following answer.

If for all $x \geq 0$, the integral v given by (1) satisfies

$$\frac{1}{2}(1 - e^{-ax^2})^{\frac{1}{2}} \leq v \leq \frac{1}{2}(1 - e^{-bx^2})^{\frac{1}{2}}, \tag{3}$$

then it is necessary and sufficient that $0 \leq a \leq \frac{1}{2}$ and $b \geq 2/\pi$.

The proof of this statement is simple. First,

$$\lim_{x \to 0} v^2/(1 - e^{-bx^2}) = (2\pi b)^{-1}.$$

Hence if (3) is true, $b \geq 2/\pi$. On the other hand (3) implies $x^{-2} \log\{1 - 4v^2\} \leq -a$ for all real x. Since the limit of this ratio, as $x \to \infty$, is 2, we have $a \leq \frac{1}{2}$. Finally

$$4v^2 = \int_{-x}^{x} \int_{-x}^{x} (2\pi)^{-1} e^{-\frac{1}{2}(s^2+t^2)} ds\,dt \geq \int_0^{2\pi} \int_0^x (2\pi)^{-1} e^{-\frac{1}{2}r^2} r\,dr\,d\theta.$$

Therefore
$$v \geq \frac{1}{2}(1 - e^{-\frac{1}{2}x^2})^{\frac{1}{2}}. \tag{4}$$

Pólya showed that as x varies from 0 to ∞, the ratio of the L.H.S. (left-hand side) of (2) to the R.H.S. decreases steadily from 1 to a minimum value and then increases steadily. Williams's calculations indicate that, approximately, the minimum value 0.9930 is taken as $x = 1.6$. Using a similar method to that of Pólya, it can be shown that the ratio of the L.H.S. of (4) to the R.H.S. is a steadily decreasing function of x for all $x \geq 0$; for the derivative of this ratio has the same sign as

$$2(e^{\frac{1}{2}x^2} - 1) - x e^{\frac{1}{2}x^2} \int_0^x e^{-\frac{1}{2}t^2} dt, \tag{5}$$

which is non-positive since
$$e^{\frac{1}{2}x^2} \int_0^x e^{-\frac{1}{2}t^2} dt = \sum_{n=1}^{\infty} \frac{x^{2n-1}}{1.3.\dots.(2n-1)}.$$

As a consequence, we obtain that this ratio (of the L.H.S. of (4) to the R.H.S) has an upper bound $2/\pi^{\frac{1}{2}}$.

2. A different lower bound for v can be obtained easily from a result proved by Chu & Hotelling (1954). There we showed that for all $x \geq 0$,

$$x^2(1 - 4v^2)(4v^2) \leq \frac{1}{2}\pi. \tag{6}$$

Hence it follows that
$$v \geq \frac{1}{2}[2x^2/(\pi + 2x^2)]^{\frac{1}{2}}. \tag{7}$$

For easy reference, we will give here a proof of (6). Let

$$g_0(x) = x^2(1 - 4v^2)/(4v^2), \tag{8}$$

then $\lim_{x \to 0} g_0(x) = \frac{1}{2}\pi$. We will show that $g_0(x)$ is decreasing. Let a prime denote differentiation with respect to x. Then

$$g_0'(x) = x(2v^3) g_1(x), \tag{9}$$

where
$$g_1(x) = v(1 - 4v^2) - x v'. \tag{10}$$

$$g_1(z) = g_0(z)\, v',$$

where

$$g_0(z) = z^3 - 12v^3,$$

$$g_0'(z) = (12/\pi)\, e^{-z^2} g_2(z),$$

where

$$g_2(z) = (\pi/6)\, z\, e^{z^2} - e^{2z^2} \int_0^z e^{-\frac{1}{2}t^2}\, dt.$$

(11)

From (5), we have
$$g_2(z) = \sum_{n=0}^{\infty} \left\{ \frac{\pi}{6 n^2} - \frac{1}{1\cdot 3 \cdots (2n+1)} \right\} z^{2n+1}.$$

It can be shown, by a similar argument used by Pólya for a similar purpose, that

$$g_2(z) = z^3(a_0 z^{-2} + a_1 + a_2 z^2 + \ldots),$$ (12)

where $a_0 < 0$ and $a_i > 0$, $i = 1, 2, \ldots$. Hence there exists an $r_1 > 0$ such that $g_2(r) < 0$ if $0 < r < r_1$, and $g_2(r) > 0$ if $r > r_1$. So as r increases from 0 to ∞, $g_2(r)$ decreases steadily from 0 to a minimum and then increases steadily to ∞. Consequently $g_2(r)$ first decreases steadily and then increases steadily. As $\lim_{r \to 0} g_1(r) = \lim_{r \to \infty} g_1(r) = 0$, it becomes clear that $g_1(r) \leqslant 0$ for all $r > 0$. Therefore $g_1(r)$ is a decreasing function of z. Hence we have (6).

Comparison can be made easily of the two lower bounds for v given by (4) and (7). For simplicity, they will be denoted by $a(x)$ and $b(x)$ respectively. Now $a(x) \gtrless b(x)$ here according as $c(x) = x^4 - 2x^2 - \pi - 1 \lessgtr 0$. As x varies from 0 to ∞, $c'(x)$, the derivative of $c(x)$, changes sign from negative to positive. So does $c(x)$. If $x = x_0$ is the solution of $c(x) = 0$, then $x_0 = 1$ approximately (the exact value is slightly smaller). Therefore, the lower bound in (7) is closer to v than that in (4) if $0 \leqslant x \leqslant 1$ (approximately) and less close if $x > 1$.

Further, the following statement is of similar nature to the one made in § 1.

If for all $x \geqslant 0$,
$$v > \tfrac{1}{2}[ax^2/(1 + ax^2)]^{\frac{1}{2}},$$ (13)

then it is necessary and sufficient that $0 \leqslant a \leqslant 2/\pi$. On the other hand, for no finite a can the R.H.S. of (13) be, for all $x \geqslant 0$, an upper bound for v.

The above statement can be shown easily by considering the limit, as $x \to 0$, of the ratio of v to the R.H.S. of (13); and the limit, as $x \to \infty$, of $(1 - 4v^2)(1 + ax^2)$.

3. Several authors have derived inequalities for Mills's ratio. Their results can be written in the form of bounds for the normal integral. For example, in our notation, Gordon's (1941) inequalities are equivalent to

$$\frac{1}{2} - \frac{1}{x}(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2} \leqslant v \leqslant \frac{1}{2} - \frac{x}{x^2+1}(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2} \quad \text{for} \quad x > 0.$$ (14)

Birnbaum (1942) improved Gordon's upper bound in (14) and obtained

$$v \leqslant \frac{1}{2} - \frac{(4 + x^2)^{\frac{1}{2}} - x}{2}(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2} \quad \text{for} \quad x \geqslant 0.$$ (15)

More recently, Tate (1953) showed what amounts to

$$\left(\frac{1}{4} + \frac{e^{-x^2}}{2\pi x^2}\right)^{\frac{1}{2}} - \frac{e^{-\frac{1}{2}x^2}}{x(2\pi)^{\frac{1}{2}}} \leqslant v \leqslant \tfrac{1}{2}(1 - e^{-x^2}) \quad \text{for} \quad x \geqslant 0.$$ (16)

We will now compare briefly (2) and (4) with (14), (15) and (16). The upper bound in (16) is obviously not so good as that in (2). The lower bound in (16) is non-negative for all real x. It is $\gtrless$ the R.H.S. of (4) according as $h(x) = x^3 - (8/\pi)(1 - e^{-\frac{1}{2}x^2}) \lessgtr 0$, as will be seen by squaring the difference twice. As x varies from 0 to ∞, $h(x)$ decreases steadily from 0 to a minimum and then increases steadily to ∞; and vanishes at $x = 1.01$ approximately (the exact value is slightly smaller). Therefore the lower bound in (16) is closer to v than that in (4) if and only if $x \geqslant 1.01$ approximately.

The lower bound in (14) is an increasing function of x for all $x > 0$. It is non-negative when $x \geqslant 0.65$ (approximately); and in this case it is $\lessgtr$ the R.H.S. of (4) according as $g(x) = 2(2/\pi)^{\frac{1}{2}} x - x^2 - (2/\pi) e^{-\frac{1}{2}x^2} \lessgtr 0$. As x varies from 0 to ∞, $g(x)$ increases steadily from $-2/\pi$ to a maximum and then decreases steadily to $-\infty$. The two roots of $g(x) = 0$ are approximately $x = 0.5$ and $x = 1.45$. Hence the lower bound in (14) is closer to v than that of (4) if and only if $x \geqslant 1.45$ approximately.

Finally we conclude that that, for topics of value, and the approximation is adequate to judge from two or three percent for large values of q, but that it may be better. No general statement is collected yet.

I am glad to acknowledge a formal that I derive for explaining processes to R. L. Lewis, some and suggestions relating to the required least mean inspection, if the are particular tunes consults. Thanks are also due of typhoid A. and H. Irving for the critical loading of the manuscripts.

REFERENCES

BIRNBAUM, Z. W. (1942). *Ann. Math. Statist.* **13**, 245–6.

GNEDENKO, D. V. & HOFFDING, H. (1944). The properties of the sample median. To be published.

GORDON, R. D. (1941). *Ann. Math. Statist.* **12**, 364–6.

KOLMOGOROV, A. (1949). *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, pp. 63–78. Berkeley: University of California Press.

TATE, R. F. (1953). *Ann. Math. Statist.* **24**, 132–4.

WILLIAMS, J. D. (1946). *Ann. Math. Statist.* **17**, 363–8.

Substitutes for χ^2

By J. B. S. HALDANE

Department of Biometry, University College, London

Neyman (1930) and Jeffreys (1948, p. 176) have suggested a substitute for χ^2 involving some saving of computation. I here suggest what I believe to be a better one. If a sample consists of N individuals belonging to m classes, and n_r belong to the rth class, the expected number on some hypothesis being

Na_r, where $\sum_{r=1}^{m} a_r = 1$, then

$$\chi^2 = \sum_{r=1}^{m} \frac{(n_r - Na_r)^2}{Na_r}.$$

Neyman's

$$\chi'^2 = \sum_{r=1}^{m} \frac{(n_r - Na_r)^2}{n_r}.$$

I consider

$$\chi''^2 = \sum_{r=1}^{m} \frac{(n_r - Na_r)^2}{n_r + 2}. \tag{1}$$

Since there is a finite probability that any n_r should be zero, it is clear that the expectation of χ'^2 is formally infinite. I shall show that it still exceeds $m - 1$ even when samples in which any $n_r = 0$ are excluded. Haldane (1953) gave reasons for preferring $n_r + 1$ as a divisor in a similar context. It can be shown that

$$\mathcal{E}\left[\sum_r \frac{(n_r - Na_r)^2 + b}{n_r + c}\right] = m - 1 + N^{-1}\{(b - c + 2)\sum a_r^{-1} - (3 - c)m + 1\} + O(N^{-2}).$$

Hence to avoid an infinite expectation c must be positive, and to avoid a multiple of Σa_r^{-1}, which may be large, in the expectation, we must have $b = c - 2$. The value $b = 0$ gives a simple formula, though $b = 1$ gives an expectation nearer to $\mathcal{E}(\chi^2)$ when N is large.

Let $n_r = Na_r + x_r$. Then

$$\chi^2 = N^{-1}\sum_r x_r^2 a_r^{-1},$$

$$\chi'^2 = N^{-1}\sum_r x_r^2 a_r^{-1}\left(1 + \frac{x_r}{Na_r}\right)^{-1}$$

$$= \chi^2 + \sum_{i=2}^{\infty} [N^{-i}\sum_r (-x_r)^{i-1} a_r^{-i}],$$

$$\chi''^2 = N^{-1}\sum_r x_r^2 a_r^{-1}\left(1 + \frac{x_r + 2}{Na_r}\right)^{-1}$$

$$= \chi^2 + \sum_{i=1}^{\infty} [N^{-i-1}\sum_r x_r^2(-x_r - 2)^i a_r^{-i-1}].$$

To find the expectations of these quantities we require the expectations of powers of x_r, namely,

$$\mathscr{E}(x_r) = 0, \quad \mathscr{E}(x_r^2) = Na_r(1-a_r), \quad \mathscr{E}(x_r^3) = Na_r(1-a_r)(1-2a_r), \quad \mathscr{E}(x_r^4) = 3N^2a_r^2(1-a_r)^2 + O(N).$$

If we write $\mathscr{E}^*(x_r^i)$ to mean the expected value of x_r^i when n_r is not zero, we omit the cases where $x_r = -Na_r$, which have a probability $(1-a_r)^N$, which tends to zero quicker than any negative power of N. Thus

$$\mathscr{E}^*(x_r) = \frac{Na_r(1-a_r)^N}{1-(1-a_r)^N}, \quad \mathscr{E}^*(x_r^2) = \frac{Na_r(1-a_r)\,1 - N^2a_r^2(1-a_r)^N}{1-(1-a_r)^N}, \quad \text{etc.}$$

So
$$\begin{aligned}
\mathscr{E}(\chi^2) &= N^{-1}\sum_{r=1}^{m}(1-a_r) = m-1, \\
\mathscr{E}(\chi'^2) &= \infty, \\
\mathscr{E}^*(\chi'^2) &= m-1+N^{-1}(2\textstyle\sum a_r^{-1}-3m+1)+O(N^{-2}), \\
\mathscr{E}(\chi''^2) &= (m-1)\left(1-\frac{1}{N}\right)+O(N^{-2}).
\end{aligned} \tag{2}$$

Thus even if we exclude the samples where any n_r is zero, χ'^2 has a positive bias often exceeding twice the reciprocal of the smallest expectation. The bias of χ''^2 is smaller, and readily calculated. The higher moments of the distribution of χ''^2 and of χ'^2, provided samples where any $n_r = 0$ are excluded, differ from those of χ^2 by quantities of the order N^{-1}. Errors of this order are neglected in the ordinary use of χ^2, and can be neglected in that of χ''^2, since χ^2 would be used if great precision were required.

As a numerical example, suppose that the numbers expected in four classes are 63, 21, 21 and 7, those observed being 71, 13, 16 and 12. Then $\chi^2 = 8\cdot825$, $\chi'^2 = 9\cdot470$, $\chi''^2 = 8\cdot319$. If we reverse the signs of the deviations, so that the observed numbers are 55, 29, 26 and 2, we find $\chi^2 = 8\cdot825$, $\chi'^2 = 16\cdot832$, $\chi''^2 = 10\cdot330$. The addition of the bias $0\cdot0268$ to χ''^2 gives values of $8\cdot345$ and $10\cdot357$, and this correction is clearly negligible. It is clear that χ''^2 is a far better approximation than χ'^2, and as it is no harder to calculate, it should be preferred.

REFERENCES

HALDANE, J. B. S. (1953). A class of efficient estimates of a parameter. *Bull. Int. Statist. Inst.* **33**, 231–48.
JEFFREYS, H. (1948). *Theory of Probability.* Oxford University Press.
NEYMAN, J. (1930). *Bull. Int. Statist. Inst.* **24**, 44–86.

A problem in the significance of small numbers

By J. B. S. HALDANE

Department of Biometry, University College, London

Dr H. Grüneberg and Dr A. G. Searle have each presented me with the following problem. N mice were examined anatomically. Before examination they had been classified into m groups, containing $n_1, n_2, n_3, \ldots, n_r, \ldots, n_m$ members. The members of a group were in fact grouped together on the basis of common ancestry. They could have been grouped on the basis of phenotypic resemblance. In one and only one of these groups, containing n members, a mice were found with a specific anatomical abnormality. What is the probability that such a coincidence should occur as the result of random sampling?

This question, like all such questions, is rather vaguely stated; and the answer to it may depend on the theory of probability adopted by the answerer. The answer here given is therefore not the only possible one.

We can arrange a fourfold table as follows:

	Normal	Abnormal
One group	$n-a$	a
Other groups	$N-n$	0

The application of Fisher's (1935) 'exact' method gives

$$P_1 = \frac{(N-a)!\,n!}{N!\,(n-a)!}$$

as the probability that all the abnormals should belong to this particular group. If they had all belonged to any one group this would have attracted notice provided P_1 was fairly small. If, however, one group had consisted of about half the total, that is to say $n = \frac{1}{2}N$ approximately, and a had not exceeded 3, the question of significance would not have arisen. The question can therefore be put somewhat more concretely: 'What is the probability that if there are a or more abnormals, all should occur in a group with n or less members?' Or still more definitely: 'It there were just a abnormals, and each group consisted of n members, what is the probability that by an accident of sampling, all the abnormals should be found in one group?'

The probability that all the abnormals should be found in one group is

$$P_2 = \frac{(N-a)!}{N!} \sum_{r=1}^{m} \frac{n_r!}{(n_r-a)!}.$$

If each group consists of just n members, which implies that N/n is an integer, this probability is

$$P_3 = P_1 \times \frac{N}{n} = \frac{N}{n}\frac{(N-a)!\,n!}{N!\,(n-a)!} = \frac{(N-a)!\,(n-1)!}{(N-1)!\,(n-a)!}.$$

I suggest that this is a reasonable estimate of the probability of the observed event, or of one equally or more unlikely, even if the values of n_r are unequal. We see that if $a = 1$, $P_3 = 1$, which is reasonable, since it is certain that the one abnormal will fall into one of the groups. We can then frame the question as follows: 'The first abnormal individual was found in a certain group. What is the probability that the $a-1$ abnormals found among the other $N-1$ mice should also be members of this group?' Clearly the value found, i.e. P_3, is a reasonable answer to this question.

In the case propounded to me by Dr Searle, $N = 472$, $n = 23$, $a = 2$, $m = 4$. So

$$P_3 = \frac{470!\,22!}{471!\,21!} = \frac{22}{471} = 0.047.$$

The uncorrected value of P is $P_1 = 0.0023$. At least one of the other groups must have consisted of 150 or more mice. Had the two abnormals belonged to this group no question of significance would have arisen. The estimation of P as about 0.05 rather than about 0.002 allows for the fact that, if the coincidence in question is an effect of random sampling, a number of other comparable coincidences would not be considered significant.

Any test of significance is somewhat arbitrary. For example, in place of the χ^2 test where

$$\chi^2 = \sum_{r=1}^{n} \frac{[a_r - \mathscr{E}(a_r)]^2}{\mathscr{E}(a_r)},$$

where a_r is an observed number and $\mathscr{E}(a_r)$ its expectation, we could use

$$\sum_{r=1}^{n} \frac{|a_r - \mathscr{E}(a_r)|}{[\mathscr{E}(a_r)]^{\frac{1}{2}}},$$

or many other similar statistics. χ^2 is used because it has a fairly simple distribution. On the same data the alternative might give a higher or a lower value of P. The test here suggested is easy to apply, and answers a reasonable question.

I think, however, that the problem merits a fuller discussion, and that a solution based on a different approach might be of equal or greater value.

REFERENCE

Fisher, R. A. (1935). The logic of inductive inference. *J. R. Statist. Soc.* **98**, 39–54.

Bounds for the ratio of range to standard deviation

By GEORGE W. THOMSON

Ethyl Corporation, Detroit 20, Michigan

1. This note supplements the work of David, Hartley & Pearson (1954) on the distribution of the ratio of the range w to the standard deviation estimate s, where both are from the *same sample* of size n. Bounds are shown to exist for w/s for all populations with non-zero variance and percentage points are given for samples of three from a normal population.

2. The bounded nature of the distribution of w/s has not been noted by any of the authors who have investigated this statistic*. It can be readily shown that the upper and lower bounds for w/s for samples from *any* population with non-zero variance arise from certain simple configurations of sample points. The upper bound, which corresponds to minimum s for a given range w, results from the arrangement with $n-2$ of the points at the sample mean and the other two points at equal distances from the mean. The lower bound, which corresponds to maximum s for a given w, results from the concentration of half

Table 1. *Bounds of the distribution of the ratio of range to standard deviation,*
w/s, in samples of size n from a normal population

n	Percentage points		n	Percentage points		n	Percentage points	
	Lower 0%	Upper 0%		Lower 0%	Upper 0%		Lower 0%	Upper 0%
			10	1·897	4·243	30	1·966	7·616
			11	1·915	4·472	40	1·975	8·832
			12	1·915	4·690	50	1·980	9·899
3	1·732	2·000	13	1·927	4·899	60	1·983	10·863
4	1·732	2·449	14	1·927	5·099	80	1·987	12·570
						100	1·990	14·071
5	1·826	2·828	15	1·936	5·292			
6	1·826	3·162	16	1·936	5·477	150	1·993	17·263
7	1·871	3·464	17	1·944	5·657	200	1·995	19·950
8	1·871	3·742	18	1·944	5·831	500	1·998	31·591
9	1·897	4·000	19	1·949	6·000	1000	1·999	44·699
			20	1·949	6·164			

of the sample points at one extreme and the other half (plus one, if the sample size is odd) of the sample points at the other extreme. The numerical values of the bounds can be shown to be:

Upper bound of w/s: $\sqrt{\{2(n-1)\}}$.

Lower bound of w/s: $2\sqrt{\{(n-1)/n\}}$ for n even,

$2\sqrt{\{n/(n+1)\}}$ for n odd.

As n becomes larger the lower bound approaches 2. It is also evident that these bounds are distribution-free provided that the points can be distributed at all. Table 1 shows the numerical values which correspond to the lower and upper 0 % points in Table 6 of the paper by David *et al.*

* [*Editorial Note*. The existence of these limits has no doubt been noticed by others who have considered the distribution of this ratio; in the correspondence leading to the joint paper by David, Hartley & Pearson (1954), the first author gave these limits in a letter of February 1954, but they were omitted in the published paper. E.S.P.]

3. Certain properties of the distribution of samples of size 3 from a normal population have been obtained by Lieblein (1952), as noted by Seth (1950). It is easily shown from Lieblein's results that the percentage points of w/s for $n = 3$ are given by

$$2 \cos [30° (1 - F)],$$

where F is the cumulative frequency. Thus, for the upper 10% point, $F = 0.90$, $w = 2 \cos 3 = 1.99726$. Table 2 shows the upper and lower 0·0, 0·5, 1·0, 2·5, 5·0 and 10·0% points and the median. The upper percentage points agree to the third decimal place with those given by David *et al.* (1954, Table 6).

Table 2. *Percentage points of the distribution of the ratio of range to standard*
deviation, w/s, in samples of size 3 from a normal population

Lower percentage points	w/s	Upper percentage points	w/s
0·0	1·73205	10·0	1·99726
0·5	1·73466	5·0	1·99931
1·0	1·73726	2·5	1·99983
2·5	1·74499	1·0	1·99997
5·0	1·75763	0·5	1·99999
10·0	1·78201	0·0	2·00000
Median			
50·0	1·93185		

4. The bounds for w's have been used in our research laboratory for the past twelve years in routine checks of the computation of s for small samples, since gross errors can be detected at once.

REFERENCES

DAVID, H. A., HARTLEY, H. O. & PEARSON, E. S. (1954). *Biometrika*, **41**, 482–93.
LIEBLEIN, J. (1952). *J. Res. Nat. Bur. Stand.* **48**, 255–68.
SETH, G. R. (1950). *Ann. Math. Statist.* **21**, 298–301.

On the estimation of population parameters from marked members

By J. A. GULLAND

Fisheries Laboratory, Lowestoft

In commercial fish populations there exist two distinct types of mortality, that due directly to capture by man, and that due to other causes. For simplification it is not unreasonable to assume that these may be represented by the constant exponential coefficients, denoted by F and M. It is the separation and estimation of these which forms a major part of any study of such populations.

Typically, the total population is of the order of millions, and recaptures from a marking experiment consist of all the marked fish caught by any fisherman and recaptures may therefore be considered as a continuous process rather than at discrete intervals, as in most experimental trapping (e.g. Hammersley, 1953). If a known number of marked individuals is released, then direct maximum likelihood estimates for the two mortalities F and M (in the marked population) may be obtained.

For the probability of an individual remaining alive until time t, when released at time $t = 0$, is $e^{-(F+M)t}$. Hence the probability of recapture in interval $(t, t + dt)$ is $F e^{-(F+M)t} dt$. Suppose now N individuals are released, of which n are recaptured at time $t_1, t_2, ..., t_n$, and that these are the only recaptures up to time T ($T > $ all t), when the experiment was ended. Then the probability of being recaptured is

$$\int_0^T F e^{-(F+M)t} dt = \frac{F}{F+M}(1 - e^{-(F+M)T}),$$

and the probability of not being recaptured is

$$\frac{M}{F+M}+\frac{F}{F+M}e^{-(F+M)T}.$$

The likelihood function may therefore be written

$$e^{L} = \binom{N}{n}\prod_{i=1}^{n}\left(Fe^{-(F+M)t_i}\right)\left(\frac{M}{F+M}+\frac{F}{F+M}e^{-(F+M)T}\right)^{N-n}.$$

Hence
$$L = \log\binom{N}{n}+n\log F-(F+M)\Sigma t_i+(N-n)\{\log(M+Fe^{-(F-M)T})-\log(F+M)\}.$$

$$\frac{\partial L}{\partial F} = \frac{n}{F}-\Sigma t_i+(N-n)\left\{\frac{(1-FT)e^{-(F+M)T}}{M+Fe^{-(F-M)T}}-\frac{1}{F+M}\right\}$$

and
$$\frac{\partial L}{\partial M} = -\Sigma t_i+(N-n)\left\{\frac{1-FTe^{-(F+M)T}}{M+Fe^{-(F-M)T}}-\frac{1}{F+M}\right\}.$$

If the experiment is continued so long that $e^{-(F-M)T}$ may be neglected, then these equations become very simple, namely,

$$\frac{\partial L}{\partial F} = \frac{n}{F}-\Sigma t_i-\frac{N-n}{F+M},\quad \frac{\partial L}{\partial M} = -\Sigma t_i+\frac{(N-n)F}{(F+M)M}.$$

Putting these equal to zero gives the solution

$$\hat{F} = \frac{n^2}{N\Sigma t_i},\quad \hat{M} = \frac{(N-n)n}{N\Sigma t_i}.$$

These estimates are biased, in fact having infinite expectations and higher moments. However, $\Sigma t_i/n$ is an unbiased estimate of $(F+M)^{-1}$, when $e^{-(F+M)T}$ is neglected, and in practice, for reasonably large values of n and N the distribution of $\hat{F}$ and $\hat{M}$ for repeated experiments is likely to be quite reasonable.

REFERENCE

HAMMERSLEY, J. M. (1953). *Biometrika*, **40**, 265–78.

REVIEWS

Biomathematics. By CEDRIC A. B. SMITH. London: Charles Griffin and Co. Ltd. Pp. xv + 712. 80s. net.

Biomathematics, by the late Dr W. M. Feldman, appeared in 1923 and reached a second edition in 1935. The present book is modestly described by Dr Smith as a third edition. But it has been entirely rewritten and is, in effect, a new book.

Although advanced statistical theory uses nearly all branches of mathematics, except perhaps projective and differential geometry, it draws on them to very varying extents. The student of statistics can go a long way with only a nodding acquaintance with analytic geometry and differential equations, but finds a knowledge of calculus and advanced algebra almost essential from the outset. He therefore requires a special course of mathematics for statisticians and will not find it within the covers of any single book. There is a great and growing need for a work of this kind.

Dr Smith's book does for biologists what one would like to see done for statisticians in general. It provides a thorough grounding in the ideas and techniques necessary for those who want to be mathematical biologists without going so far as to be biological mathematicians; and it does so with great fluency and insight without any sacrifice of rigour. There can be few writers as well qualified as Dr Smith to write such a work. His mathematical powers, his extensive knowledge and practical experience of mathematical and statistical applications in biology, the care he has bestowed on the work and his gifts as a teacher combine to make this a work of outstanding excellence. It will be useful not only to biologists but to any student who requires a competence in mathematics for the purposes of pursuing his own subject.

The book opens with two chapters on arithmetic, including some calculating devices and some account of mechanical equipment and punched card machines. Chapter 3 is a refresher course on algebra, leading to the treatment of inequalities in Chapter 4. Chapter 5 deals with the connexion between algebra and geometry and provides a foundation in analytic geometry and conic sections. The next two chapters deal with logarithms (which Dr Smith introduces by means of the equi-angular spiral) and their applications in computation by slide-rules and nomograms. Chapters 8–12 are on differential and integral calculus and proceed as far as simple differential equations and partial differentials. Chapter 13 deals with series and Chapter 14 with vectors. Methods of solving equations and matrices receive separate chapters, and the book closes with three chapters on Chance, Statistical Distributions and Simple Statistical Procedures, together with one on Colson's method of simplifying arithmetical calculations. The book contains over 700 pages and the amount of ground covered is astonishing considering that no impression is given of hurry or over-consideration.

Readers of Stephen Leacock will remember his protests about the dullness of arithmetical examples, and his attempt to dispel it by investing with human attributes those three anonymous characters of our youth *A*, *B* and *C*; especially the case of poor *C*, who died of pneumonia contracted while trying to fill a cistern with a leak in it. Dr Smith, who combines with his other qualities a sense of humour, would have endeared himself greatly to Leacock. For instance, the additive properties of matrices are illustrated on the population of Narkover, the determination of minima is exemplified, not on those eternal salmon tins, but on the angles of origin of human arteries; the treatment of physical magnitudes is enlivened by an example leading to the conclusion that if one falls into a cold sea the best thing to do is to swim hard so as to keep warm.

Altogether, this is a very sound, readable and sensible book. In spite of its price it should become widely popular.

M. G. KENDALL

Introduction to Mathematical Statistics. 2nd edition. By PAUL G. HOEL. New York: John Wiley and Sons Inc.; London: Chapman and Hall. 1954. Pp. ix + 331. 40s.

This is a much expanded and revised version of a book that first appeared in 1947. Apart from some expansion of many topics such as regression, the analysis of variance and non-parametric methods, there have been notable additions in the probability field. As a result the book is much more self-contained than formerly. It presupposes a knowledge of calculus—probably about two years study and seems designed for scientists who have some mathematical background and wish to master the

elements of statistical theory. For the applications to their own fields, they will have to turn to other texts.

Inevitably in a book of this size some things have had to be omitted, for example, time series, equalization of variance and probit analysis. It is perhaps surprising, in view of the intended public, that no mention is made of Sheppard's corrections for grouping or the many short-cut procedures available nowadays, such as the range in the analysis of variance. There is also an unusual differentiation between large sample and small sample distributions — is the F distribution a small sample distribution? Diagrams are used excellently throughout the book and are well labelled. There are numerous exercises for which no answers are provided. The book ends with a number of useful tables, but whilst the percentage points of t, F and χ^2 are given, those for the normal curve have to be obtained by inverse interpolation in the table of the cumulative distribution provided.

<div style="text-align: right">P. G. MOORE</div>

Design and Analysis of Industrial Experiments. Edited by O. L. DAVIES. London and Edinburgh: Oliver and Boyd, for Imperial Chemical Industries Limited. 1954. Pp. xiii + 636. 63s.

This book is written by a team of authors largely the same as those who wrote *Statistical Methods in Research and Production* in 1947, but, as explained in the Introduction, the present volume deals with the design of experiments and their subsequent analysis rather than the extraction of information from previously existing data. The particular point of the book is that it is written from a broadly chemical point of view, rather than the more usual agricultural one, and so the design of experiments is freed from the restrictions imposed by agricultural conditions, and sometimes inadvertently carried over into fields where they do not apply.

Successive chapters deal with Simple Comparisons, Sequential Tests—a remarkably fine chapter —Sampling and Testing Methods, Randomized Blocks, and Incomplete Randomized Blocks, again very good because it concentrates attention on the simpler designs required for industrial experimentation. It is nice to have the assurance that these are worth while, for the alternative view that in them too few observations have to bear a heavy load of theoretical interpretation is somewhat prevalent. Then there are four chapters on Factorial Experiments and one on the determination of optimum conditions. In this last chapter the treatment is largely orographical, which I like as it enables simple words like 'ridge' to be used as short-cuts avoiding much difficult mathematical explanation. Then follows a glossary which is far the worst part of the book. No readers of a book of this type need to have explained what is quaintly called "The arithmetic average or mean', and if they do the definition here given will hardly help them, nor is it possible to explain 'Universe, Population, Parameter, Sample, Statistic' in fourteen lines. The final tables are well arranged.

The book as a whole contains a very large amount of valuable information, but as might be expected from a team the style varies greatly in clearness and readability; certainly the best parts are very good. There is an implied assumption that nothing but the experiment and its own conditions need be considered in planning, and this is certainly not always true for those who work in less magnificent concerns than I.C.I., but still you must know how the experiment really ought to be done, before planning a compromise with what can be done, and the necessary information is here.

Finally, it must be emphasized that in industry an investigation is not complete when the results have been obtained in conventional statistical form. It is necessary to translate them into the language used by the executive who is to decide what action to take on them. It is avowedly not the purpose of this book to deal with this question, but there is internal evidence that at least one of the authors would be well able to do so, and I should like to suggest that in a future edition such a chapter should be added in place of the glossary.

<div style="text-align: right">L. McMULLEN</div>

Sample Survey Methods and Theory. By M. H. HANSEN, W. N. HURWITZ and W. G. MADOW. New York: John Wiley and Sons Inc.; London: Chapman and Hall. 1953. Vol. I, Methods and Applications. Pp. xxii + 638. 64s. Vol. II, Theory. Pp. xiii + 332. 56s.

These books aim to cover the whole field of the sample survey and cater for all tastes and all classes of statistician. Vol. I sets out the methods in common use in sampling surveys and the way in which the methods are applied. It begins with commonsense talk about the fundamental aims of sampling and sampling design and continues with the delineation of such statistical ideas as may be expected to be

useful to the particular purpose of the authors. The chapter on bias and non-sampling errors in survey results is obviously written from practical experience. Most of the book is concerned with descriptions of the various types of sampling—simple random sampling, stratified simple random sampling, simple one- or two-stage cluster sampling, stratified single or multi-stage cluster sampling. In each case the procedure is clear, every possible query which might be raised by the would-be user is answered. The chapter on estimating variances and the accuracy of the method of estimation will not be helpful unless the reader has learned a little elementary statistics but presumably if he has reached this point in the book he will willy-nilly have acquired such basic information as he will here find to be necessary.

Vol. II can either be read in conjunction with Vol. I or separately when it can be looked on as an elementary text-book exercise in statistical algebra. The standard here is uneven. It is surely unnecessary to explain 'why we study summation notation' (p. 11) to someone who is expected to understand without explanation why we study convergence in probability (p. 72). None the less in spite of irrelevancies the algebraic framework is clear and concise and should not cause undue difficulty to any student interested enough to read the book. Certain contractions are possibly invented by the writers and should be rejected; 'plim' standing for 'limit in the sense of probability' is one of the least useful of these. All the formulae used in vol. I are derived here, and the way is clear for the reader to make such modifications as his own particular sampling design may dictate.

The two books taken together cover fully the whole field of sample surveys, and are unlikely to be superseded for some years to come. This being the case it is unfortunate that such encyclopaedic works should be priced out of the possible range for many persons.

F. N. DAVID

Sampling Theory of Surveys with Applications. By Pandurang V. Sukhatme. New Delhi: The Indian Society of Agricultural Statistics; Ames, Iowa: The Iowa State College Press. 1954. Pp. xxviii + 491. $6.

This is the fifth book dealing specifically with sample survey methods and theory to appear in the last five years. The author is head of the statistics branch of F.A.O. and has wide experience of sample surveys in India and elsewhere, and one might have expected his book to be more closely concerned than it is with the methods of overcoming the difficulties arising in agricultural surveys and in surveys of underdeveloped territories. In fact, Dr Sukhatme has written a text-book on the algebraic development of the standard branches of sample-survey theory. On this basis the treatment is painstaking and clear and the book will be most valuable to the student or research worker who requires an ad hoc treatment of the theory appropriate to any particular sampling method.

The up-to-date character of the book is shown by the prominence given to unequal-probability sampling, the treatment of which is based in part on the author's own research, and by the chapter on non-sampling errors, which contains some novel results on the treatment of interviewing errors.

However, it is to be regretted that Dr Sukhatme in common with other recent writers on sample survey theory has neglected the opportunity to base his treatment of the theory on the simplifying principles set out by Yates five years ago in *Sampling Method for Censuses and Surveys*. Writing mainly for practical workers rather than for theoreticians—no proofs of the formulae were given—Yates based his exposition on the fact that most survey designs are built up of the following components:

(a) Method of selection, e.g. simple random, systematic, probability proportional to size.
(b) Stratification.
(c) Use of supplementary information.
(d) Multi-stage sampling.
(e) Multi-phase sampling.

The number of combinations of the various possibilities under each of these heads is so large that it would be impossible to give within reasonable bounds the appropriate estimation and variance formulae for all sample designs likely to be useful in practice. However, Yates pointed out that the contribution from each component could be considered separately and the results combined to give the relevant formulae for any particular design.

As an example, the estimate of a mean based on a stratified sample normally takes the form $y = \Sigma N_i y_i / \Sigma N_i$, where y_i is the estimate of the ith stratum mean. Suppose that by referring to the theory of unstratified sampling under (a) above, we were able to calculate an estimate s_i^2 of $V(y_i)$. Yates would immediately write down the estimate of $V(y)$ as $\Sigma N_i^2 s_i^2 / (\Sigma N_i)^2$, and would assert that this result held generally whatever methods are used for sampling within strata. Sukhatme, however, has preferred to start from the beginning for each type of stratified sample and to develop explicit formulae

from first principles for each case, rather than prove a general rule for all types of stratified sample. Similarly, in his treatment of both the ratio method and multi-stage sampling, the author has preferred to treat each case separately from first principles rather than to develop the general rules by which the estimates can be built up. This is all the more regrettable since the general laws are easier to prove than the formulae for special cases.

A further opportunity for simplification was missed in the treatment of sampling with unequal probabilities with replacement. If we draw a sample of n with replacement from a population of values $y_1, y_2, ..., y_N$ with selection probabilities $p_1, p_2, ..., p_N$, all necessary estimation and variance formulae can be obtained by regarding the sample as a simple random sample from the discrete probability distribution in which the observation y_i has associated with it the probability p_i. Thus, once the theory of sampling from an infinite population with equal probabilities of selection has been worked out the corresponding results for unequal-probability sampling with replacement emerge as corollaries. In Dr Sukhatme's treatment, however, these results are worked out afresh.

For these reasons the book contains a great deal of avoidable algebra. Although some students undoubtedly find it helpful to have an ad hoc theory set out for each particular type of sample design, the consequent details are likely to prove unnecessarily time-consuming for the reader whose object is to grasp the principles of sample-survey theory.

J. DURBIN

Life and other Contingencies, Vol. I. By P. F. HOOKER and L. H. LONGLEY-COOK. Cambridge University Press. 1953. Pp. viii + 312. 22s. 6d.

This book is one of a series commissioned by the Institute of Actuaries and the Faculty of Actuaries to provide a course of reading suitable for the examinations conducted by these bodies. The present work replaces Spurgeon's *Life Contingencies*, which now ends an honourable career of some forty years. The authors have evidently been influenced to a considerable extent by their distinguished predecessor. The main changes arise from the demands of the new actuarial syllabus and consist of the introduction of sections on extra risks, valuation methods and non-mortality (sickness, maternity, etc.) benefits. There are also a number of additions necessitated by developments in theory and practice, such as the treatment of family income benefits and the appendix on International Actuarial Notation.

However, the book is by no means merely a rather thorough revision of the earlier text-book. It is a completely new work and develops the subject in its own way. Emphasis is definitely on the provision of methods for practical application in life-office work. There is little preoccupation with complicated algebraic manipulation, and the theoretical study of mortality laws and stationary populations is reduced to a minimum. Worked examples are collected at the ends of chapters and, perhaps for this reason, appear to be less profuse than in Spurgeon's text-book. There are no exercises to be worked out by the student.

The treatment of multiple decrement tables is left to vol. II, and it will be interesting to see how the authors tackle this still somewhat controversial topic. As a result of this division of subject-matter joint-life assurances and annuities are not considered in the present volume.

It is to be hoped that 'Hooker and Longley-Cook' will be as long-lived as 'Spurgeon'. If this is so the reviewer hopes to see in the successive revisions a much fuller treatment of non-mortality benefits, and, if possible, either an earlier and more adequate discussion of the construction of tables, or the omission of this subject by its transfer to another part of the actuarial syllabus.

N. L. JOHNSON

Population Statistics and their Compilation. By HUGH H. WOLFENDEN. The University of Chicago Press, for the Society of Actuaries. 1954. Pp. xxii + 258. 56s. 6d.

This book is a completely revised edition of a work originally published as 'Actuarial Study No. 3' by the Actuarial Society of America in 1925. The choice of subject-matter clearly reflects the actuarial interests of the author. In the analytical section of the book there is much greater emphasis on the study of mortality than on any other aspect of population phenomena. No fewer than 70 pages are devoted to the discussion of various methods of constructing mortality-tables and life-tables, while a further 23 pages are concerned with the comparison of mortality by occupations, causes, etc. and a brief discussion on forecasting mortality rates. There are only eight pages on the use of data on marriages, births, orphanhood & unemployment and three pages on 'Sickness Data'. Emigration and immigration are not specifically considered at all. The theory of reproductivity is relatively adequately

treated in 16 pages containing some sensible warnings on the uncritical use of the various indices of reproductivity.

The earlier part of the book shows much less evidence of this lack of balance. There is an interesting and useful description of the compilation of population statistics by means of censuses, and registration of births, marriages and deaths. This is accompanied by a discussion of likely sources of unreliability in the data, and methods of correcting the consequent defects.

Comprehensiveness has evidently been the author's ideal. While he has succeeded in making his book a mine of information on particular topics, it is to be expected that students may experience difficulty in sorting out really important material from items now mainly of historical interest. The dust-cover claims that 'this book is the only presentation, by an actuary, of the particular actuarial viewpoints and methods necessary to the production of modern population statistics'. Apart from the fact that this statement ignores books published by the Institute of Actuaries in recent years, the catalogue-like assembly of methods presented makes difficult the disentanglement of the more modern methods.

Despite these criticisms, the book should prove useful as a comprehensive, yet handy, work of reference. There is no index, though there is a detailed 'Table of Contents' by paragraphs (not by pages). There are nearly one hundred footnotes which bear witness to the author's devotion to the ideal of comprehensiveness. Most of them contain useful information, though they tend to interrupt the textual line of argument.

There is a 22-page appendix on 'Some Theory in the Sampling of Human Populations', by W. E. Deming. It is well written but will be of little value except to those with sufficient statistical training and such persons will probably be familiar with the subject already.   N. L. JOHNSON

Table of Binomial Coefficients. Royal Society Mathematical Tables. 3. London: Cambridge University Press. 1954. Pp. viii + 162. 35s.

These tables, prepared under the editorship of J. C. P. Miller, were computed mainly at the National Physical Laboratory, Liverpool University, and the Royal Aircraft Establishment at Farnborough. They give ^nC_r for

(i) $r \leqslant \frac{1}{2}n \leqslant 100$,
(ii) $2 \leqslant r \leqslant 12$, $200 \leqslant n \leqslant 500$,
(iii) $2 \leqslant r \leqslant 11$, $500 \leqslant n \leqslant 1000$,
(iv) $2 \leqslant r \leqslant 5$, $1000 \leqslant n \leqslant 2000$,
(v) $r = 2, 3$, $2000 \leqslant n \leqslant 5000$.

The computations throughout are performed by use of the recurrence relation

$$^{n+1}C_r = {^nC_r} + {^nC_{r-1}}.$$

The main application of the table lies in the field of number theory for such investigations as to the possible uses of tetrahedral numbers to form other numbers. For statisticians the combinatorial numbers mainly conjure up visions of binomial probabilities and here, as the probabilities have to be combined with the combinatorial number, the volume is unlikely to supplant such well-worn friends as the incomplete beta function, at any rate for the lower values of n. Occasions do, however, arise where binomial coefficients alone are required—certain probability problems of a combinatorial nature for example—and these tables will form a useful reference work for such cases. The volume is well set out with very clear type, and a useful index is provided to aid one in locating the value required.   P. G. MOORE

Bessel Functions and Formulae. Compiled by W. G. BICKLEY. London: Cambridge University Press, for the Royal Society. 1953. Pp. 11. 3s. 6d.

This collection of formulae is a straight reprint of the *Summary of Notations* and the section on 'Functions and Formulae' from the British Association Mathematical Tables, Vol. x, Bessel Functions, II.

It gives a summary of the differential, integral and difference relations obeyed by the Bessel and auxiliary functions and also some connecting formulae with related functions (e.g., Hankel's, Whittaker's, etc.). It further gives a variety of series expansions for and involving Bessel functions.

Of particular interest to statisticians are the sections giving generating functions and Laplace transforms of Bessel functions.   D. E. BARTON

PUBLICATIONS OF U.S. DEPARTMENT OF COMMERCE, NATIONAL BUREAU OF STANDARDS

(i) Tables of Chebyshev Polynomials $S_n(x)$ and $C_n(x)$. Applied Mathematics Series, 9. 1952. Pp. xxix + 161. $1.75.

Defining the nth order polynomials as

$$C_n(x) = 2\cos\{\operatorname{arc\,cos} \tfrac{1}{2}x\}, \quad S_n(x) = 2(4 - x^2)^{-\frac{1}{2}} \sin\{(n+1)\operatorname{arc\,cos} \tfrac{1}{2}x\},$$

the main tables gives their values to 12 decimal places for $m = 2(1)\,12$, $x = 0(0.001)\,2$. Subsidiary tables give the expansions of these polynomials in powers of x, together with those of the modified forms:

$$T_n(x) = \tfrac{1}{2}C_n(2x), \quad T_n^*(x) = \tfrac{1}{2}C_n(4x - 2),$$
$$U_n(x) = S_n(2x), \quad U_n^*(x) = S_n(4x - 2).$$

Also given are inverse relations for powers of x as linear sums of $T_n(x)$ and $T_n^*(x)$ up to the twelfth power.

These polynomials are chiefly of use as ancillaries to the computing of functions possessing power series expansions over a whole range of values of the argument, as the convergence of a series of Chebyshev polynomials is more uniform and in general more rapid than that of a power series expansion taken about a point of the range. The computation was carried out under the direction of A. N. Lowan and a 22-page introduction is written by Cornelius Lanczos. This gives, *inter alia*, the example of the asymptotic expansion of the incomplete normal integral, where the Chebyshev series to six terms gives one-fifth the error of approximation of the power series expansion of the same order, for the possibility of deviating more than $\sqrt{2}$ standard deviations from the mean. These polynomials are distinct from those commonly used in statistical practice and also described as Chebyshev's, defined by

$$P_m(x) = \Delta^m\left[\binom{x + \tfrac{1}{2}(n-1)}{m}\binom{x - \tfrac{1}{2}(n+1)}{m}\right].$$

(ii) Tables of coefficients for the numerical calculation of Laplace transforms. Applied Mathematics Series, 30. 1953. Pp. 36. 25 cents.

These tables give coefficients for the approximate evaluation by quadrature of the Laplace transform

$$F(p) = \int_0^\infty f(t)\, e^{-pt}\, dt$$

of a function $f(t)$ which is given for n equally spaced values of t whose range includes the effective range of the argument. The coefficients are given to nine decimal places for $n = 2(1)\,11$ and varying ranges and intervals of p depending on n (e.g. $p = 0.1(0.1)\,1.0$ for $n = 2$, $p = 1(1)\,10$ for $n = 11$). Special tables are also given for the simpler case where $f(t)$ is a low order polynomial. A short introduction is provided by H. E. Salzer. In this he remarks: 'Convenient estimates of the error of approximation seem difficult to obtain.' However, in the example given of $f(t) = J_0(t)$ it is found that for $n = 11$, $F(p)$ differs from the approximating function by less than $0.14\,\%$ over the range of p considered.

D. E. BARTON

iii) Tables of Lagrangian Coefficients for sexagesimal interpolation. Applied Mathematics Series, 35. 1954. Pp. ix + 157. $2.00.

These tables give 3-, 4-, 5- and 6-point Lagrangian interpolation coefficients A, for arguments in sexagesimal measure, such as angles given in units of degrees, minutes and seconds. The coefficients are given for 3600 values of the fraction of the tabular interval. Thus if the function is tabled for each degree, the coefficients may be used to find a value of the function at any required minute and second. Each coefficient is tabled to eight decimal places. There is a brief Introduction by H. E. Salzer.

(iv) **Tables of circular and hyperbolic sines and cosines for radian arguments.** Applied Mathematics Series. 36. 1953. Pp x + 407. $3.00.

The main Table I gives for $x = 0(0.0001)1.9999$ values to nine decimal places of the functions sin x, cos x, sinh x and cosh x. There are three short supplementary tables:

Table II gives values to nine decimals of the same four functions for $x = 0.0 \, 0.1 \, 10.0$
Table III is a conversion table for expressing degrees, minutes and seconds in radians and vice versa.
Table IV gives to 15 decimal places values of $n \times \frac{1}{2}\pi$ for $n = 1(1)100$.

There is a brief Introduction by A. N. Lowan.

(v) **Table of secants and cosecants to nine significant figures at hundredths of a degree.** Applied Mathematics Series. 40. 1954. Pp. vi + 46. 35 cents.

This tables gives sec x and cosec x for $x = 0.00(0.01)90.00$ degrees. It will serve as a companion to the table of sin x and cos x to fifteen decimal places at hundredths of a degree previously published as No. 5 in the Applied Mathematics Series.

E. S. PEARSON

CORRIGENDA

Biometrika (1954), **41**

(1) M. E. WISE, p. 328. For equation (8·5) read:
$$Nx = (N - \tfrac{1}{2}n + \tfrac{1}{2})(1 - h) + \tfrac{1}{2}(p - 1)$$

(2) P. WHITTLE, p. 437. On the left-hand side of equation (16)
for $\xi_t - a\xi_{t-1} - b\xi_{t-2}$ *read* $\xi_t - a\xi_{t-1} - b\xi_{t+1}$

(3) H. A. DAVID, p. 466. In Table 2, the result for the rectangular population with $n = 4$
for 1·019 *read* 1·010

(4) D. R. COX, p. 472. In Table 2, for population (*b*), the value for $n = 4$:
for 1·961 *read* 1·939, and for $n = 5$: *for* 2·252 *read* 2·196

The Corrigenda to papers by Rushton and Ruben stated in the List of Contents for Vol. **41**, Parts 3 and 4, as printed on p. 568, were wrongly placed *in front* of the first page (p. 287) of that issue.

Tables of the Incomplete B-Function

Edited by KARL PEARSON

59 pages of Introduction and 494 pages of Tables

Price: 55s. net

Tables of the Incomplete Γ-Function

Edited by KARL PEARSON

31 pages of Introduction and 164 pages of Tables

Price: 42s. net

Tables of the Complete and Incomplete Elliptic Integrals

from LEGENDRE's *Traité des Fonctions Elliptiques*. With autographed portrait of LEGENDRE

39 pages of Introduction by KARL PEARSON and 94 pages of Tables

Price: 12s. 6d. net

Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples

By F. N. DAVID

38 pages of Introduction, 55 pages of Tables, 10 Diagrams and 4 Charts

Price: 17s. 6d. net

NEW PUBLICATION NOW AVAILABLE

Biometrika Tables for Statisticians, Vol. I

The two volumes of *Tables for Statisticians and Biometricians* are now out of print and will not be re-issued.

At the request of the Biometrika Trustees a complete recasting of these Tables has been undertaken by Professor E. S. PEARSON and Dr H. O. HARTLEY. Many of the old tables have been set aside or modified, tables which have been published during the last fifteen years in *Biometrika* are reproduced and some new tables added.

Volume I of the new series, which includes the statistical and auxiliary mathematical tables in more common use is now available. It contains an Introduction and 54 tables covering in all 238 pp.

Price: 25s. net

NEW STATISTICAL TABLES: SEPARATES RE-ISSUED FROM BIOMETRIKA

To be obtained from

BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON, W.C.1

I. From *Biometrika*, Vols. 22, 27 and 28

   Tests of Normality. By E. S. PEARSON and R. C. GEARY

   Price Two Shillings and Sixpence, post free

II. From *Biometrika*, Vol. 32, Part 2, pp. 168–181 and 188–189

   (1) **Table of percentage points of the incomplete beta-function**
   (2) **Table of percentage points of the χ^2 distribution**

   Stitched together with introductory matter. Price Two Shillings and Sixpence, post free

III. From *Biometrika*, Vol. 32, Parts 3 and 4, pp. 300–310

   (1) **Table of the probability integral of the range in samples from a normal population**
   (2) **Table of the percentage points of the range**
   (3) **Table of the percentage points of the t-distribution**

   Stitched together with introductory matter. Price Two Shillings and Sixpence, post free

IV. From *Biometrika*, Vol. 33, Part 1, pp. 73–88

   Table of percentage points of the inverted beta (F) distribution

   With introductory matter. Price Two Shillings and Sixpence, post free

V. From *Biometrika*, Vol. 33, Part 3, pp. 252–265

   (1) **Table of the probability integral of the mean deviation in samples from a normal population**
   (2) **Table of the percentage points of the mean deviation**

   Stitched together with introductory matter. Price Two Shillings and Sixpence, post free

VI. From *Biometrika*, Vol. 33, Part 4, pp. 296–304

   Table for testing the homogeneity of a set of estimated variances

   With introductory matter. Price Two Shillings, post free

VII. From *Biometrika*, Vol. 35, Parts 1 and 2, pp. 145–156

   Table of significance levels for the Fisher-Yates test of significance in 2×2 contingency tables. By D. J. FINNEY

   With introductory matter. Price Two Shillings and Sixpence, post free

VIII. From *Biometrika*, Vol. 35, Parts 1 and 2, pp. 191–201

   Table for the calculation of working probits and weights in probit analysis. By D. J. FINNEY and W. L. STEVENS

   With introductory matter. Price Two Shillings and Sixpence, post free

IX. From *Biometrika*, Vol. 36, Parts 3 and 4, pp. 267–289

   Tables of autoregressive series. By M. G. KENDALL

   With introductory matter. Price Two Shillings and Sixpence, post free

X. From *Biometrika*, Vol. 36, Parts 3 and 4, pp. 431–449

   Tables of symmetric functions, Part I. By F. N. DAVID and M. G. KENDALL

   With introductory matter. Price Two Shillings and Sixpence, post free

NEW STATISTICAL TABLES: continued

Biometrika Index

A Biometrika Index comprising Subject Index for Volumes 1–37 and Author Index for Volumes 1–40 is now available

Price: 6s. net or $1.00

To be obtained from

BIOMETRIKA OFFICE, UNIVERSITY COLLEGE, LONDON, W.C.1

BIOMETRICS

Journal of the Biometric Society

TRABAJOS DE ESTADÍSTICA

REVIEW PUBLISHED BY *INSTITUTO DE INVESTIGACIONES ESTADÍSTICAS*
OF THE *CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS*

MADRID, SPAIN

CONTENTS

Vol. VI   *Cuad. I*

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

CONTENTS

VOL. 50, NO. 271   **SEPTEMBER 1955**

POPULATION ESTIMATION BASED ON CHANGE OF COMPOSITION
CAUSED BY A SELECTIVE REMOVAL

By DOUGLAS G. CHAPMAN*

*Department of Mathematics, University of Washington and Lectureship in the
Design and Analysis of Scientific Experiment, Oxford University*

1. INTRODUCTION

It has been noted in wildlife studies that it is possible to estimate the size of an animal
population on the basis of the change of sex ratio following a kill concentrated on one sex.
It is difficult to ascertain to whom this idea should be credited, and what name should be
given to it. Scattergood (1954), in a general survey, lists several references to the method.
Chapman (1954) called the procedure the dichotomy method of estimation and showed that
the usual 'intuitive' estimates are also maximum likelihood estimates. It is not clear that
this title is sufficiently descriptive, and in any case the classification does not necessarily
need to be dichotomous. On the other hand, change of composition does seem to reflect the
essential aspect of the procedure.

The assumptions involved in this method appear to be simpler and more likely to be
fulfilled than those underlying other population estimation procedures. Moreover, it will
be shown that non-fulfilment of some of the assumptions may involve no serious con-
sequences. This is one of the phases of the method studied in this paper.

To be balanced against this is the fact that the amount of information yielded is smaller
for the same effort than is obtainable from the more usual method of estimation based on
the recapture of marked members of the population. A quantitative comparison is given
below of the two methods. Some problems of optimum design in the sampling that accom-
panies a change of composition estimate are considered and some extensions of the pro-
cedure studied.

2. BINOMIAL MODEL

The simplest situation arises if a 'closed' population is assumed, i.e. one where there is
no emigration or immigration and births and deaths are negligible over the period of the
experiment—except for the removal process. As to the removal process, all that is required
of it is that the final or cumulative total be known and a breakdown into at least two classes
or components. The third assumption for the simple binomial model is that random samples
have been taken of the population before and after the removal process. The randomness
is in respect to the specified classes or components.

The following notation is needed:

Unknown parameters: $\begin{cases} N_i = \text{population size at time } t_i \ (i = 0, 1) \text{ made up of two} \\ \qquad \text{classes } X \text{ and } Y, \\ X_i, Y_i = \text{size of classes } X \text{ and } Y \text{ at times } t_i. \end{cases}$

It is convenient to write $P_i = X_i/N_i$ $(i = 0, 1)$:

Known parameters:
$$\begin{vmatrix} n_i = \text{size of random samples taken at times } t_i, \\ R_x = X_1 - X_0 = \text{removal from class } X, \\ R_y = Y_1 - Y_0 = \text{removal from class } Y, \\ R = R_x + R_y = \text{total removal.} \end{vmatrix}$$

Random variables: $x_i, y_i =$ the number of classes X and Y respectively in sample i $(i = 0, 1)$.

Analogous to P_i, we define $p_i = x_i/n_i$ $(i = 0, 1)$.

As noted in Chapman (1954) the maximum-likelihood estimates of X_0 and N_0 are

$$\hat{X}_0 = \frac{x_0(n_1 R_x - x_1 R)}{n_1 x_0 - n_0 x_1} = \frac{p_0(R_x - p_1 R)}{p_0 - p_1}, \tag{1}$$

$$\hat{N}_0 = \frac{n_0(n_1 R_x - x_1 R)}{n_1 x_0 - n_0 x_1} = \frac{R_x - p_1 R}{p_0 - p_1}, \tag{2}$$

while the asymptotic variances of X_0 and N_0 are

$$\sigma^2(\hat{X}_0) = \frac{\dfrac{X_0 Y_0}{n_0} P_1^2 + \dfrac{X_1 Y_1}{n_1} P_0^2}{(P_0 - P_1)^2}, \tag{3}$$

$$\sigma^2(\hat{N}_0) = \frac{\dfrac{X_0 Y_0}{n_0} + \dfrac{X_1 Y_1}{n_1}}{(P_0 - P_1)^2}. \tag{4}$$

In the usage made of this method in wildlife studies it is N_0 that is ordinarily estimated, based as noted above, on the change of sex ratio. The sports kill records show the removal of each sex. In many cases the kill is confined to males; this represents a very favourable situation. It should be pointed out that under the assumptions made it is immaterial whether attention is focused on time t_0 or time t_1.

In the possible application of this procedure to the estimation of fish populations, it must be recognized the removal is not always selective by sex. However, here the classification may be by size or by species. In many fisheries there is a size restriction on fish that may be retained by the fishery. If the biologist can sample in a manner to include randomly other sizes, then this change of composition estimation procedure may be used. In this case, or in case the classification is by species, the experimenter will usually be more interested in estimating X_0.

It may be suggested that where X and Y represent qualitatively different populations, one should expect different catchabilities and consequently the procedure based on such a classification would be unsatisfactory. However, in the rather important case that $R_y = 0$, the estimate remains asymptotically unbiased even if the catchability of the X population is not the same as that of the Y population (for the experimenter or sampler). We repeat that the catchabilities associated with the removal process are irrelevant.

For consider that X and Y have catchabilities λ_x and λ_y and assume the total sample catch has a Poisson distribution. Then given n_i, the sample catch size, x_i, has a binomial distribution with parameters n_i, $X_i/(N_i - Y_i\delta)$, where $\delta = 1 - \lambda_y/\lambda_x$.

Then replacing the random variables in $\hat{X}_0$ by their expectations (which they will converge to in probability as the sample sizes are increased) we have

$$\hat{X}_0 = \frac{X_0 N R_x - X_x R - \delta Y_x R - \delta R_x R_y}{N_0 R_x - X_0 R - \delta Y_x R - \delta N_0 R_y}$$

and the result stated follows immediately.

However, $\hat{X}_0$ does not have the same desirable property, for replacing x, y, by their expectation under the model of unequal catchabilities

$$\hat{X}_0 = (N_0 - Y_0 \delta)\frac{(N_0 R_x - X_0 R - \delta Y_x R_x + \delta R_x R_y)}{N_0 R_x - X_0 R - \delta Y_x R - \delta N_0 R_y}$$ (b)

Nevertheless, the fact that X_0 is still estimable indicates that it may be possible to use a quite different species in this estimation procedure e.g. a scrap fish may form the Y class while the desirable sports fish is the X class. It should be pointed out that the distribution of the estimate will be modified, in case $\lambda_x \ne \lambda_y$ even though $R_y = 0$. In particular, in this case the asymptotic variance of $\hat{X}_0$ is

$$\sigma^2(\hat{X}_0(\delta)) = \frac{1-\delta}{(P_0 - P_1)^2}\left[\frac{X_0 Y_0}{n_0}\frac{P_1^2}{(1-P_0\delta)^2} + \frac{X_1 Y_1}{n_1}\frac{P_0^2}{(1-P_1\delta)^2}\right].$$ (7)

The variation of $\sigma^2(\hat{X}_0)$ with respect to δ depends on all of the unknown parameters. However, it is possible to state some useful qualitative results. These can be determined simply from a study of the function

$$f(\delta) = (1-\delta)/(1-P_i\delta)^2 \quad (\delta < 1),$$

which attains its maximum at $\delta = 2 - 1/P_i$. This maximum is $[4P_i(1-P_i)]^{-1} > 1$.

The qualitative results are summarized in the following table:

	$\delta > 0$ $\lambda_x > \lambda_y$	$\delta < 0$ $\lambda_x < \lambda_y$
$P_0, P_1 < \frac{1}{2}$	$\sigma^2(\hat{X}_0(\delta)) < \sigma^2(\hat{X}_0(0))$	$\sigma^2(\hat{X}_0(\delta)) > \sigma^2(\hat{X}_0(0))$
	$\sigma^2(\hat{X}_0(\delta)) > \sigma^2(\hat{X}_0(0))$	$\sigma^2(\hat{X}_0(\delta)) < \sigma^2(\hat{X}_0(0))$
$P_0, P_1 > \frac{1}{2}$		

The favourable cases occur in the main diagonal; they are favourable in the sense that if $\sigma^2(\hat{X}_0(0))$ is used to estimate the variance or determine a confidence interval, any errors incurred will be on the conservative side. In terms of a working rule for the sample considered, if the undesirable fish are abundant relative to the desirable fish, then the procedure is more satisfactory if the sampling device is such that the desirable fish are more catchable. Since $f(\delta)$ achieves a maximum value of unity at $\delta = 0$ for $P_i = \frac{1}{2}$, it seems reasonable to suggest that when the P_i straddle $\frac{1}{2}$ and do not differ greatly from it, $\sigma^2(\hat{X}_0(\delta))$ will differ little from $\sigma^2(\hat{X}_0(0))$.

One other possibility of error in this procedure arises from the fact that R_x, R_y and R may not be true totals removed from the population. In general, it is to be expected that the removal will be under-evaluated because of the unreported kill as well as the unknown natural mortality.

Suppose that the actual removals are $R_x + \Delta_x$, $R_y + \Delta_y$, $R + \Delta$, where $\Delta_x + \Delta_y = \Delta$ and where we may write $\Delta_x = KR_x$, $\Delta = KR + \epsilon$.

Then for $n_i \to \infty$, the estimate $\hat{X}_0$ converges in probability to

$$X_0 \left(\frac{N_0 R_x - X_0 R - \epsilon R_x}{(N_0 R_x - X_0 R)(1 + K) - \epsilon X_0} \right). \tag{8}$$

It is convenient to denote the term in the brackets of (8) by $1 + b$. X_0 also converges in probability, under these conditions, to $X_0(1 + b)$.

Consider first the case where $\epsilon = 0$, i.e. the unknown removal of X and Y is in the same proportion as the known removal. Then

$$1 + b = (1 + K)^{-1} \approx 1 - K \text{ for } K \text{ small};$$

hence the relative asymptotic bias is negative and of the order of the unknown proportion of the removal.

For the more important case where ϵ is not zero we restrict attention to the situation where $R_y = 0$. Then

$$1 + b = \frac{Y_0 - \epsilon}{Y_0(1 + K) - \frac{\epsilon X_0}{R}}.$$

If the mortality or unknown removal of the X's and Y's is proportional to the size of these groups, then

$$\Delta_y = Y_0 \Delta_x / X_0 = Y_0 KR / X_0$$

and

$$\Delta = KRN_0 / X_0, \quad \epsilon = Y_0 KR / X_0.$$

Hence

$$b = - KR / X_0.$$

In this situation the relative bias is still negative and of an even smaller order than in the case where $\epsilon = 0$.

In the usage of (4) (or (3)) to estimate the large sample variances, in a situation where part of the removal is unknown to the experimenter, errors will arise in the incorrect estimation of the X_i, Y_i. It is easily seen that, while on the average, for sufficiently large samples, X_0, Y_0 are underestimated, X_1 and Y_1 will be overestimated and consequently the estimation errors will compensate. From a qualitative point of view, therefore, it appears that the errors due to bias in the estimation of the variances are of much smaller consequence than the sampling errors.

The cases considered above are by no means the only possibilities that might occur with a portion of the removal unknown to the experimenter. They represent, however, two cases of practical importance. Of course if no assumptions whatsoever are made as to the size and relative magnitudes of Δ_x and Δ_y, it is clear that they may be such as to vitiate the estimation procedure.

3. Optimum sample allocation

Since this method of population estimation involves two samples over which the experimenter has some degree of control, it is natural to inquire as to the optimum disposition of effort between these samples. If N_0 is to be estimated with $\sigma^2(\hat{N}_0)$ to be minimized, subject to the restriction that $n_1 + n_0$ be fixed, then elementary calculus leads to the following equation for the ratio of n_1 to n_0:

$$\frac{n_1}{n_0} = \left(\frac{X_1 Y_1}{X_0 Y_0} \right)^{\frac{1}{2}}. \tag{9}$$

Formula (9) involves parameters which are unknown at the outset of the experiment. It is possible to use it to suggest qualitative rules of procedure. Consider the most favourable case where the removal is restricted entirely to one class, say the X class, i.e. $R_y = 0$. Then

$$n_1 = n_0 \left(1 - \frac{R}{X_0}\right)^{\frac{1}{2}};$$

the ratio n_1/n_0 is given in Table 1 for various values of R/X_0, the proportion of the X class removed. Since the removal will rarely exceed 50 % of the removed class, this table suggests that if N_0 alone is to be estimated, n_1 should be chosen only slightly smaller than n_0.

Table 1. *Optimum sample ratio n_1/n_0 for the estimation of N in case there is no removal in the Y class*

R/X_0	0·5	0·4	0·3	0·2	0·1	0·05	0·02	0·01
n_1/n_0	0·71	0·77	0·84	0·89	0·95	0·975	0·990	0·995

If X_0 is to be estimated, the situation is slightly more complicated, but a qualitative rule is possible. Recalling $\sigma^2(X_0)$ from formula (3), this is minimized with $n_0 + n_1$ fixed if

$$\frac{n_1}{n_0} = \left(\frac{X_1 Y_1}{X_0 Y_0}\right)^{\frac{1}{2}} \frac{P_0}{P_1} = \frac{N_1}{N_0} \left(\frac{X_0 Y_1}{X_1 Y_0}\right)^{\frac{1}{2}}. \tag{10}$$

Again, consider the case $Y_0 = Y_1$ and hence $R_r = R$. Then the optimum choice of n_0 and n_1 is determined by

$$\frac{n_1}{n_0} = \left(1 - \frac{R}{N_0}\right)\left(1 - \frac{R}{X_0}\right)^{-\frac{1}{2}},$$

which is shown in Table 2.

Table 2. *Optimum sample ratio n_1/n_0 for the estimate of X in case there is no removal in the Y class*

$P_0 = X_0/N_0$	$R N_0 = r$						
	0·25	0·20	0·15	0·10	0·05	0·02	0·01
0·1	—	—	—	—	1·344	1·096	1·043
0·2	—	—	1·700	1·273	1·097	1·033	1·015
0·3	1·838	1·386	1·202	1·102	1·041	1·014	1·007
0·4	1·225	1·132	1·075	1·039	1·016	1·005	1·003
0·5	1·061	1·032	1·016	1·007	1·001	1·000	1·000
0·6	0·982	0·980	0·982	0·986	0·993	0·997	0·998
0·7	0·935	0·947	0·959	0·972	0·985	0·994	0·997
0·8	0·905	0·924	0·943	0·963	0·981	0·993	0·996
0·9	0·882	0·907	0·931	0·954	0·977	0·991	0·996

Table 2 suggests that in a few extreme cases the second sample should be considerably larger than the first for an optimum estimate of X_0; for most cases, however, the samples should be chosen almost equal in size. Since often the biologist will wish to know both X_0 and N_0, it would seem that choosing $n_0 = n_1$ represents a near optimum allocation of effort.

4. Comparison of Capture-Recapture and Change
of Composition Estimation Procedures

Rather than merely sampling and classifying as X or Y $n_0 + n_1$ animals, the experimenter may tag or mark an equivalent number. The removal process will then act as a sample to estimate the tag ratio. It is perhaps reasonable to disregard the additional effort necessary to record accurately the tag recoveries. However, it is not reasonable to disregard the fact that tagging may be an operation that is more expensive, or requires more effort than classification.

In view of the result of the last section it is reasonable to consider the case $n_1 = n_2 = n$ (say). For simplicity we restrict consideration to $R_y = 0$, $R_x = R$. Denote R/N_0 by r. Then

$$\sigma^2(\hat{N}_0) = \frac{Y_0}{n}\left[\frac{X_0 + X_0 - R}{(1-P_0)^2}\right]\left(\frac{1-r}{r}\right)^2$$

$$= \frac{N_0^2}{n}\left[\frac{2P_0 - r}{1 - P_0}\right]\left[\frac{1-r}{r}\right]^2. \tag{11}$$

Now consider the estimate that could be made if $2\lambda n$ tags were placed in the population, where λ^{-1} is the ratio of the cost of tagging to the cost of classification and hence may be assumed to be greater than or equal to 1.

The random tag recoveries follow a hypergeometric distribution under the usual assumption of random sampling without replacement; this hypergeometric distribution is often approximated by a Poisson distribution (e.g. Chapman, 1948). Here, if n is considered to be small relative to N, while R is not necessarily so, then it may be shown by the usual procedures (using Stirling's formula for the large factorials appearing in the hypergeometric formula) that the distribution tends in the limit, as $N \to \infty$, while $n/N \to 0$ and R/N is bounded away from zero, to the binomial distribution with parameters $(n, R/N)$. If this limiting distribution is used as an approximation to the true hypergeometric distribution, the asymptotic variance of the estimate of N_0 based on the tag recoveries (say $\hat{N}_0(t)$) is

$$\sigma^2(\hat{N}_0(t)) = \frac{N_0^2(N_0 - R)}{2\lambda n R} = \frac{N_0^2(1 - r)}{2\lambda n r}. \tag{12}$$

Hence we may ask under what conditions on λ, r and P_0 is

$$\frac{1}{2\lambda} > \left(\frac{2P_0 - r}{1 - P_0}\right)\left(\frac{1-r}{r}\right). \tag{13}$$

This inequality holds provided

$$P_0 < \frac{r + 2r\lambda(1-r)}{4\lambda(1-r) + r}. \tag{14}$$

Let

$$f(r, \lambda) = \frac{r + 2r\lambda(1-r)}{4\lambda(1-r) + r}. \tag{15}$$

We observe that $f(r, 0) = 1$, while $f(r, 1)$ is a monotone function of r with $f(0, 1) = 0$. $f(1, 1) = 1$. The first statement in this sentence is the trivial remark that the change of composition method must always be better if tagging is prohibitively costly. The second case, where tagging is no more costly than classification, is of some interest; we consider this case in the next paragraph.

Elementary considerations show that for $r < 1$, $P_0 = f(r-1)$ implies $P_0 < r$. Since the Y removal is zero, it is impossible for the proportion of the population removed to exceed the original proportion of X's. Consequently, if tagging is no more costly than classification, from the large sample point of view, the tag sample procedure is under all circumstances more efficient than the change of composition estimation method.

Turning to the case $\lambda > 1$, we have used a different method of comparison. Tabled below is the minimum value of 1λ such that $\sigma^2(\dot N_0(t)) > \sigma^2(\dot N_0)$ for specified values of P_0.

Table 3. *Minimum value of relative cost of tagging to cost of classification such that* $\sigma^2(\dot N_0(t)) > \sigma^2(\dot N_0)$

| $X_i N_0 = P_0$ | $R N_0 = r$ | | | | | | |
	0·25	0·20	0·15	0·10	0·05	0·02	0·01
0·1	—	—	—	-	6·34	19·60	41·80
0·2	—	—	3·54	6·76	16·62	46·55	96·53
0·3	3·00	4·58	7·28	12·86	29·85	81·20	166·90
0·4	5·51	8·00	12·28	21·00	47·49	127·40	260·70
0·5	9·00	12·80	19·27	32·40	71·20	192·08	392·04
0·6	14·25	20·00	29·75	49·50	109·25	289·10	589·05
0·7	23·00	32·00	47·22	78·00	171·00	450·80	917·33
0·8	40·51	56·00	81·16	135·00	294·51	774·20	1574·10
0·9	93·00	128·00	87·00	306·00	332·51	1744·40	3544·20

Since the comparison of these two procedures is made under the conditions most favourable to the change of composition technique, these results suggest that in almost all cases the capture-recapture estimation procedure will yield more information for the same amount of effort. However, it is to be noted that the capture-recapture estimates are seriously dependent on the assumption that the initial capture does not alter the behaviour pattern of the animals in any way that will affect their probability of recapture. This assumption is difficult to test and it is certainly not always reasonable.

5. Estimation when R_x and R_y must be estimated

In the field, many additional complications may arise in the application of the change of composition estimation procedure. Often R is known reasonably well, but R_x and R_y are not; it may even happen that R itself must be estimated. These complications do not change the estimates but will modify the estimated variances and any interval estimates.

We study the case where R is known but R_x and R_y must be estimated from a sample of size m from the removed group. We assume that a sample of size m is taken randomly, and that of the m animals, m_x are from class X, m_y from Y. Then the estimates of X_0 and N_0, denoted by $\tilde X_0$, $\dot N_0$, are simply (1) and (2) with R_x replaced by $\hat R_x$, where

$$\hat R_x = \frac{m_x R}{m}. \tag{16}$$

While the asymptotic variances of X_0 and N_0 may be determined by inverting a 3×3 information matrix, it is easier to compute them directly, for $\bar{X}_0$ and $\bar{N}_0$ are linear functions of R_x. Thus

$$\sigma^2(\bar{N}_0) = (P_0 - P_1)^{-2} \left[\left(\frac{X_0 Y_0}{n_0} + \frac{X_1 Y_1}{n_1} \right) + \frac{m R_x R_y}{R^2 (P_0 - P_1)^2} \left\{ 1 + 3 \frac{P_0(1-P_0)}{n_0} + 3 \frac{P_1(1-P_1)}{n_1} \right\} \right]. \quad (17)$$

This is the asymptotic variance, asymptotic in the sense that terms of higher order in n_0^{-1}, n_1^{-1} are neglected.

A similar formula can be derived for $\sigma^2(\bar{X}_0)$, and the same method can be followed through in the case that R itself has to be estimated in some manner. It is of course assumed that the estimation of R is independent of the random variables x_0, x_1. The assumption is implicit in the development above.

6. Combination of capture-recapture and change of composition estimates

It was noted in §4 that, if the experimenter is prepared to accept the assumptions underlying the capture-recapture estimates, then he will gain more from putting his effort into tagging animals rather than merely classifying them. Actually, however, he may try to do both, for in this way he may get two different estimates, or the combined procedure may yield the most efficient single estimate. Moreover, the biologist will often wish to sample the population on more than one occasion, in order to study growth or other changes in the population. Consequently if the members of the initial sample are tagged as the basis of a tag-sample estimate, the normal pattern of the study of the population will make possible also the change of composition estimate (assuming of course a selective removal between the samples).

In this situation the first problem that confronts the experimenter is to determine whether there is reasonable agreement between the two possible estimates. As noted above in §3, the binomial model with parameters $(n_0, R/N_0)$ is a reasonable approximation to represent the capture-recapture procedure. Since the large-sample variances are known for the change of composition estimate and for the estimate based on the number of tags recovered (cf. formula (12)), a large-sample test is available to compare the two separate estimates.

If the estimates are not compatible it may be due to failure in the assumptions in one or both models. These assumptions may be correct in regard to expectations, but the sources of variability may be greater than the models permit, i.e. the distributions are not binomial. This possibility will be considered in the next section.

On the other hand, there may be sources of error, such as were discussed in §2, or are well known for capture-recapture estimates. These have been particularly well pointed out by De Lury (1954). It is pertinent to note that the most important sources of error tend to bias the change of composition estimate downward, while the important factors of trap-shyness and tag mortality will tend to inflate the tag sample estimates upwards.

If the estimates are compatible, there is a suggestion but certainly no proof that the assumptions of the two models are fulfilled. In this case it will be normal practice to combine them. They may be averaged, weighting them inversely to their asymptotic variances. However, for this simple model the maximum-likelihood equations are easily solved. We need to add to the notation of §2 the following:

$s =$ number of the n_0 animals taken in sample 0

which are recovered in the removal process.

It is assumed of course that these n_0 animals have been made identifiable by tagging or marking in some manner. Then

$$p(x_0, x_1, s) = \prod_{i=0}^{1} \binom{n_i}{x_i} \left(\frac{X}{N}\right)^{x_i} \left(\frac{Y}{N}\right)^{y_i} \binom{r_s}{s} \left(\frac{R}{X_0}\right)^s \left(1 - \frac{R}{X_0}\right)^{\cdots}. \tag{18}$$

The maximum-likelihood equations for X_0 and N_0 are seen to be of the same form as those of the simple model of § 2. They may be written in the form

$$\frac{x_0}{X_0} - \frac{2n_0}{N_0} + \frac{x_1}{X_1} - \frac{n_1 - n_0 + s}{N_1} = 0, \tag{19}$$

$$\frac{y_0}{Y_0} - \frac{2n_0}{N_0} + \frac{y_1}{Y_1} - \frac{n_1 - n_0 + s}{N_1} = 0, \tag{20}$$

and solved by analogy with those of the simple model. The maximum-likelihood estimators are

$$\hat{X}_0^c = \frac{x_0[(n_1 - n_0 + s) R_x - x_1 R]}{(n_1 - n_0 + s) x_0 - 2n_0 x_1}, \tag{21}$$

$$\hat{N}_0^c = \frac{2n_0[(n_1 - n_0 + s) R_x - x_1 R]}{(n_1 - n_0 + s) x_0 - 2n_0 x_1}. \tag{22}$$

The information matrix is

$$\begin{pmatrix} \dfrac{n_0}{X_0 Y_0} + \dfrac{n_1}{X_1 Y_1}, & -\dfrac{n_0}{N_0 Y_0} - \dfrac{n_1}{N_1 Y_1} \\[2mm] -\dfrac{n_0}{N_0 Y_0} - \dfrac{n_1}{N_1 Y_1}, & \dfrac{n_0 X_0}{N_0^2 Y_0} + \dfrac{n_1 X_1}{N_1^2 Y_1} + A \end{pmatrix}, \tag{23}$$

where

$$A = \frac{n_0 R}{N_0^2 N_1}. \tag{24}$$

Hence

$$\text{(Asymptotic)} \quad \sigma^2(\hat{N}_0^c) = \left[\frac{(P_0 - P_1)^2}{\dfrac{X_0 Y_0}{n_0} + \dfrac{X_1 Y_1}{n_1}} + \frac{n_0 R}{N_0^2 N_1} \right]^{-1}. \tag{25}$$

An analysis of (25) shows that the optimum allocation of n_0 and n_1 is made by choosing n_0 the maximum possible, n_1 the minimum. This, of course, is necessarily true from the results obtained in § 4, if the differential cost of tagging is disregarded. Consequently the combined procedure outlined here will be used if there is reason to question the assumptions of the tag-sample procedure, or if further sampling after the removal process will be made for other reasons. In this case the question of optimum allocation does not enter.

7. EXTENSIONS FROM THE SIMPLE BINOMIAL MODEL

There are several immediate extensions or generalizations of the model considered in § 2. For example, there may be several samples with one removal or several removals with one pair of samples, or several samples and several removals. The first two of these extensions are trivial; the several selective removals between two random samples may simply be combined by addition and the original model is obtained. If there are several samples either before or after a single removal, they may be combined in the usual manner that samples pertaining to the binomial model are combined.

Where there are several selective removals, each followed by a random sample, a new situation arises. This is of some interest, for the sampling process might be combined with the removal process to lead to such a situation, i.e. the sampler might only return the members of the Y group to the population (though in general it is to be expected that the experimenter will prefer to return both X's and Y's marked or tagged).

The following additional notation is needed:

$$R_{x_i}: \quad \text{removal of } X \text{ population prior to } i\text{th sample.}$$

$$R_{y_i}: \quad \text{removal of } Y \text{ population prior to } i\text{th sample.}$$

$$R_i = R_{x_i} + R_{y_i} \quad (i = 0, 1, 2, ..., k).$$

Also

$$X_i = X_0 - R_{x_i}, \quad N_i = N_0 - R_i, \quad \text{etc.}$$

The maximum-likelihood equations for X_0 and N_0 may be written down simply as

$$\sum_{i=0}^{k} \frac{x_i}{X_i} = \sum_{i=0}^{k} \frac{n_i}{N_i}, \left.\begin{array}{c} \\ \\ \end{array}\right\}$$
$$\sum_{i=0}^{k} \frac{y_i}{Y_i} = \sum_{i=0}^{k} \frac{n_i}{N_i}. \tag{26}$$

These equations can be solved by iterative methods in any particular case. In case the R_{x_i}, R_{y_i} are small relative to X_0, Y_0, the following approximate solution may be employed to yield estimates without the labour of the iterative procedure. The logarithm of the likelihood function may be written

$$\ln L = K + \sum_{i=0}^{k} x_i \ln\left(P_0 - \frac{R_{x_i}}{N_0}\right) + \sum_{i=0}^{k} y_i \ln\left(1 - P_0 - \frac{R_{y_i}}{N_0}\right) - \sum_{i=0}^{k} n_i \ln\left(1 - \frac{R_i}{N_0}\right)$$

$$\doteqdot K + \sum_{i=0}^{k} x_i \ln P_0 - \sum_{i=0}^{k} \frac{x_i R_{x_i}}{N_0 P_0} + \sum_{i=0}^{k} y_i \ln(1 - P_0) - \frac{\sum y_i R_{y_i}}{(1 - P_0) N_0} + \sum_{i=0}^{k} \frac{n_i R_i}{N_0}. \tag{27}$$

so that the likelihood equations become

$$(1 - \hat{P}_0) \sum_{i=0}^{k} x_i R_{x_i} + \hat{P}_0 \sum_{i=0}^{k} y_i R_{y_i} - \hat{P}_0 (1 - \hat{P}_0) \sum_{i=0}^{k} n_i R_i = 0. \tag{28}$$

$$\frac{\sum_{i=0}^{k} x_i}{\hat{P}_0} + \frac{\sum_{i=0}^{k} x_i R_{x_i}}{\hat{N}_0 \hat{P}_0^2} - \frac{\sum_{i=0}^{k} y_i}{1 - \hat{P}_0} - \frac{\sum_{i=0}^{k} y_i R_{y_i}}{(1 - \hat{P}_0)^2 \hat{N}_0} = 0. \tag{29}$$

$\hat{P}_0$ may be determined from the quadratic equation (28) and this in turn substituted in (29). The result is a simple linear equation for $\hat{N}_0$.

The possibility mentioned earlier that the selective removal might be performed by the sampler suggests a further possibility—to set up a sequential procedure for sampling, removing the X group in the sample from the population and so on for k steps where k is determined by the actual observations. This would be analogous to the sequential procedure proposed by Goodman (1953) for the tag sample method of estimation.

Perhaps more important than any of these extensions is to consider the situation where sampling that conforms to the binomial model is impossible. Schooling and segregation by size or sex or age will certainly be frequently encountered in many natural populations, so that factors beyond the control of the experimenter make invalid the assumptions of the

binomial model. This will be true even when he is cognizant of the principles of random sampling. Since these additional factors tend to produce more variability from sample to sample than is ascribable to the binomial model it is necessary for the experimenter to replicate his samples to obtain an internal measure of this variability.

Consider then several samples n_{ij} $(i = 0, 1; j = 1, 2, \ldots, k_i)$ each composed of x_{ij} members of the X class and y_{ij} members of the Y class. To estimate X_0, N_0 a regression technique may be used that has some resemblance to De Lury's procedure for population estimates from the catch per unit of effort data (De Lury, 1947). To do so it is necessary to postulate that the random variables $p_{ij} = x_{ij}/n_{ij}$ are approximately normal with mean P and variance σ_{ij}^2.

Now define

$$Z_{0j} = X_0 - p_{0j}N_0, \qquad \left.\begin{array}{c}\\\\\end{array}\right\}$$
$$Z_{1j} = X_0 - R_x - p_{1j}(N_0 - R). \tag{30}$$

Then
$$\mathcal{E}(Z_{ij}) = 0, \quad \sigma^2(Z_{0j}) = N_0^2\sigma_{0j}^2, \quad \sigma^2(Z_{1j}) = (N_0 - R)^2\sigma_{1j}^2. \tag{31}$$

If the underlying heterogeneity is considerable it is reasonable to assume that $\sigma_{ij}^2 = \sigma^2$ for all i, j; though if the n_{ij} differ considerably some note should be taken of this, presumably following the ideas of Cochran (1942). Also for simplification the factors N_0^2, $(N_0 - R)^2$ in the variances of the Z_{ij}'s are neglected. The maximum-likelihood equations for X_0 and N_0 then reduce to a form almost identical with (1) and (2), provided the single observations there are replaced by means.

Thus letting $p_{i.} = \dfrac{\sum\limits_{j=1}^{k_i} p_{ij}}{k_i}$,

$$\hat{N}_0 = \frac{R_x - Rp_1.}{p_0. - p_1.}, \tag{32}$$

while by routine methods it is found that

$$\text{(Asymptotic) } \sigma^2(\hat{N}_0) = \frac{\sigma^2}{(P_0 - P_1)^2}\left[\frac{N_0^2}{k_0} + \frac{(N_0 - R)^2}{k_1}\right]. \tag{33}$$

It should be pointed out that (33) is a large sample variance if k_0, k_1 are large, not n_0, n_1 as was the case with respect to formulas (3) and (4).

Formula (33) may be used to estimate the magnitude of the sampling required for a given removal, or the magnitude of the removal for given sampling, to achieve an estimate with preassigned precision. Thus if $R_y = 0$, $k_0 = k_1 = k$ (say).

$$\text{(Asymptotic) } \sigma^2(\hat{N}_0) = \frac{1}{k(1 - P_1)^2}\frac{\sigma^2}{} \left[\frac{2 - 2r + r^2}{r^2}\right]. \tag{34}$$

Unfortunately, it will often be true in advance that so little is known as to the possible range of values for σ, that this formula cannot be used even to give qualitative results.

This procedure can be extended to the case of several selective removals, with a non-selective sample taken between each removal. No new theory is involved although the algebra becomes somewhat more tedious. However, such a process must necessarily take place over time where mortality and possibly emigration and immigration should be considered. It is therefore thought desirable to defer consideration of this case to a later study, where the restriction that the population is closed, is removed.

It is planned to illustrate the methods developed above in a further paper, using appropriate data concerned with fish populations.

REFERENCES

CHAPMAN, D. G. (1948). A mathematical study of confidence limits of salmon populations calculated from sample tag ratios. *Bull. Int. Pacif. Salm. Fish. Comm.* no. 2, pp. 67–85.

CHAPMAN, D. G. (1954). The estimation of biological populations. *Ann. Math. Statist.* 25, 1–15.

COCHRAN, W. G. (1942). Sampling theory when the units are of unequal sizes. *J. Amer. Statist. Ass.* 37, 199–212.

DE LURY, D. B. (1947). On the estimation of biological populations. *Biometrics*, 3, 145–67.

DE LURY, D. B. (1954). On the assumptions underlying estimates of mobile populations. *Statistics and Mathematics in Biology*, pp. 287–93. Ames, Iowa: Iowa State College Press.

GOODMAN, LEO A. (1953). Sequential sampling tagging for population size problems. *Ann. Math. Statist.* 24, 56–69.

SCATTERGOOD, L. W. (1954). Estimating fish and wildlife: a survey of methods. *Statistics and Mathematics in Biology*, pp. 273–85. Ames, Iowa: Iowa State College Press.

AN AGE-DEPENDENT BIRTH AND DEATH PROCESS

By W. A. O'N. WAUGH

Royal Military College of Science, Shrivenham

1. INTRODUCTION

Many examples of the stochastic processes known as branching or population processes have been developed. The assumptions leading to the various models have been decided partly by mathematical convenience and partly by the attempt to reproduce features controlling the development of biological or physical populations. The model we shall develop involves a continuous time parameter and an enumerable set of states, and we can compare it with several other processes of this type, e.g. the Yule–Furry birth process, Feller's birth and death process (see Feller, 1939, 1950), and the non-Markovian birth process due to D. G. Kendall (1948).

There are two motives for the development of the present model. Mathematically, it furnishes a conveniently soluble example of a non-Markovian process, in which transitions from a given state may be to one of several other states, and in this it is a generalization of Kendall's process in which transitions are always from the given state to its successor, i.e. from state number n to state number $n+1$. For purposes of application it is based on assumptions which are not proposed as ideal but which bring in features which are perhaps a step in the direction of realism. Thus it may help in the development of yet more realistic models when they are required, and in indicating both the scope and the limitations of the models constructed under an assumption which has been widely adopted in population problems; viz. that the individuals reproduce independently of one another.

Bellman & Harris (1948, 1952) have described a stochastic process similar to the one we shall examine, and we shall adopt their notation. They consider populations of particles which reproduce by fission independently of one another. They suppose that a 'generation time' is defined for these populations which is a random variable having a general distribution function $G(t)$ (which does not depend on the size of the population, in accordance with the assumption of independence just mentioned), so that $dG(t)$ expresses the probability that a particle born at time $t = 0$ ends its life in the interval $(t, t + dt)$.

They assume that with given probabilities $q_0, q_1, q_2, \ldots$ any particle whose life has just ended is replaced by $0, 1, 2, \ldots$ new particles.

Their detailed work is carried out for the pure birth process involving binary fission only:

$$q_2 = 1, \quad q_0 = q_1 = q_3 = \ldots = 0.$$

They state that their results can easily be generalized to other modes of fission. Note that when $q_0 \neq 0$ the process is a birth *and death* process. Further developments in this direction have also been obtained by Ramakrishnan (1951) and by Reid (1953).

It may be convenient to note here that a focal point of the present investigation will be the coefficient of variation of the population size, considered as a function of the time. The importance of this function in the study of populations has been emphasized previously, e.g. by Kendall (1952a).

2. AGE-DEPENDENT PROCESS INVOLVING TWO POSSIBLE FATES: DEATH OR FISSION

For clarity and because it leads to a model which may be useful in applications, we shall not adopt, in the main part of our investigation, the most general assumptions that are possible. We shall indicate at various points throughout the paper certain further generalizations that merely involve difficulties of technique and alterations in detail which might be worked out if applications seemed to demand it.

We shall suppose that at the end of its life an individual either disappears or is replaced by two new individuals (i.e. death without issue or reproduction by binary fission). These two new individuals then develop and reproduce independently of one another in the same way as the parent individual, i.e. if the parent individual split at time t to form them, then the probability that either of the two new individuals ends its life during $(t+\tau, t+\tau+d\tau)$ is $dG(\tau)$, the distribution $G(\tau)$ being fixed, and the same for all individuals of the population. In the notation of Bellman & Harris this would be equivalent to putting

$$q_0 \neq 0, \quad q_2 \neq 0, \quad q_1 = q_3 = \dots = 0,$$

but we shall make the more general assumption that these probabilities depend on the age at which the parent individual's life ends. A further generalization to admit the possibility of an individual's facing more than two alternative fates is quite simple. We shall suppose that at time $t = 0$ the population consists of just one *newly born* individual. Let us name the alternatives which face an individual of the population; risk (b) is the event that its life ends with the birth of two new individuals, risk (d) is the event that it dies without issue. We shall derive the functions $q_j(t)$ and $G(t)$ as follows. Suppose that there are two functions $\lambda(t)$ and $\nu(t)$ such that, given that an individual born at time 0 is alive at time t,

$$\Pr\{\text{individual succumbs to risk } (b) \text{ in } (t, t+dt)\} = \lambda(t)\, dt + o(dt),$$

$$\Pr\{\text{individual succumbs to risk } (d) \text{ in } (t, t+dt)\} = \nu(t)\, dt + o(dt),$$

thus $\Pr\{\text{individual's life ends in } (t, t+dt)\} = [\lambda(t) + \nu(t)]\, dt + o(dt).$

The function $[\lambda(t) + \nu(t)]$ might be called the 'force of mortality'. In the argument which follows, and which leads to the fundamental integral equation, it is convenient if $\lambda(t)$ and $\nu(t)$ are both continuous and do not vanish together at any point in $0 < t < \infty$ and are such that $\int_0^t [\lambda(u) + \nu(u)]\, du \to \infty$ as $t \to \infty$. In the latter part of the paper we assume a certain form for $\lambda(t)$ and $\nu(t)$, and it happens that these requirements of continuity, etc., are fulfilled. However, it is possible to extend the argument to cover cases in which certain discontinuities are allowed, in particular the case where an instantaneous chain reaction can occur at the moment of fission (a possibility which is mentioned by Bellman & Harris (1952)). In this case our assumption of age-dependence permits the probabilities governing the numbers of progeny at instantaneous fission to be different to those corresponding to fission after a finite life length.

From $\lambda(t)$ and $\nu(t)$ we obtain the following conditional probabilities, given that the individual's life ends in $(t, t+dt)$:

$$\Pr\{\text{life ends by risk } (b)\} \equiv q_2(t) = \frac{\lambda(t)}{\lambda(t) + \nu(t)}.$$

$$\Pr\{\text{life ends by risk } (d)\} \equiv q_0(t) \equiv \frac{\nu(t)}{\lambda(t) + \nu(t)}.$$

Let
$$G(t) \equiv 1 - \exp\left\{-\int_0^t [\lambda(u) + \nu(u)]\, du\right\}$$

= Pr {individual born at time 0 lives for a time less than or equal to t}

We can imagine an idealized population in which either the risk (d) is suspended and the population develops under risk (b) alone ($\nu(t) = 0$) or vice versa. Then

$$B(t) \equiv 1 - \exp\left\{-\int_0^t \lambda(u)\, du\right\}$$

= Pr {individual born at time 0 lives a shorter time than t, when risk (d) is suspended}.

$$D(t) \equiv 1 - \exp\left\{-\int_0^t \nu(u)\, du\right\}$$

= Pr {individual born at time 0 lives a shorter time than t, when risk (b) is suspended}.

If
$$g(t) = \frac{dG(t)}{dt}, \quad d(t) = \frac{dD(t)}{dt}, \quad b(t) = \frac{dB(t)}{dt}.$$

then
$$q_0(t) = \frac{d(t)\,[1 - B(t)]}{g(t)}$$

and
$$q_2(t) = \frac{b(t)\,[1 - D(t)]}{g(t)},$$

and also
$$1 - G(t) = [1 - B(t)][1 - D(t)].$$

3. THE INTEGRAL EQUATION FOR THE GENERATING FUNCTION OF THE POPULATION SIZE

In order to obtain the mean and variance of the population size as functions of the time we start from an integral equation for the generating function of the distribution of the population size, which is very similar to the one from which Bellman & Harris (1952) derive their results. It is possible to give a definition of a suitable sample space Ω and a function $r_t(\omega)$ over Ω which may be called the population size at time t, and by considerations of measurability to show that $\{r_t : t \geqslant 0\}$ is a well-defined stochastic process, and furthermore to derive the fundamental integral equation from this definition. However, we will content ourselves with this reference to the more rigid approach and will merely give the following intuitive argument which leads to the integral equation.

Let $p_r(t)$ be the probability that the population size is r at time t, given that the population consisted of a single newly born individual at time 0.

Let
$$\phi(z, t) \equiv \sum_{r=0}^{\infty} p_r(t)\, z^r \equiv E z^{r_t},$$

and let
$$h(z, t) \equiv \sum_{n=0}^{\infty} q_n(t)\, z^n.$$

Then the situation at time t may have come about in either of two ways.
  (i) The initial individual may still be alive.
  (ii) The initial individual's life may have ended in the interval $(u, u + du)$ where $0 \leqslant u < t$. In this case, if its life ends in fate (b), then because of the independence of the members of the population and the fact that all individuals are assumed to develop under the same probabilities as regards life length and number of progeny, we have effectively two popula-

tions developing each for a period $t - u$, from initial ancestors newly born at time u. On the other hand, if its life ends in fate (d) then the population size at time t is zero.

Hence we have the integral equation

$$\phi(z,t) = \int_0^t [q_0(u) + q_2(u)\,\phi^2(z,t-u)]\,g(u)\,du + z[1 - G(t)],$$

$$= \int_0^t h[\phi(z,t-u),u]\,g(u)\,du + z[1 - G(t)],$$

the latter form of which can easily be seen to hold when modes of fission other than binary are considered.

In terms of the functions $B(t)$, $D(t)$ and their derivatives this can be written

$$\phi(z,t) = \int_0^t [\phi(z,t-u)]^2\,b(u)\,[1 - D(u)]\,du + \int_0^t d(u)\,[1 - B(u)]\,du + z[1 - B(t)]\{1 - D(t)\}.$$

If we differentiate this equation with respect to z and then put $z = 1$ we get the following equation for the mean population size at time t,

$$\mu_1(t) = 2 \int_0^t \mu_1(t-u)\,b(u)\,\{1 - D(u)\}\,du + [1 - B(t)][1 - D(t)].$$

while if we differentiate twice with respect to z and then put $z = 1$ we get the equation

$$\mu_2(t) = 2 \int_0^t [\mu_1(t-u)]^2\,b(u)\,[1 - D(u)]\,du + 2 \int_0^t \mu_2(t-u)\,b(u)\,[1 - D(u)]\,du,$$

for the second factorial moment of the population size.

4. THE OBSERVABLE LIFE-LENGTH DISTRIBUTIONS AND THE PROBABILITY OF EXTINCTION

In experimental observations we should probably be able to separate the data on life length into those for individuals ultimately suffering fate (b) and those for individuals ultimately suffering fate (d). These observable distributions will be related to the distributions $B(t)$ and $D(t)$ in a similar manner to that in which crude rates are related to net rates in a problem of competing risks.

Let $B^*(t)$ be the probability that an individual newly born at time 0 suffers fate (b) before time t. Then $B^*(\infty)$ is the probability that a newly born individual will ultimately suffer fate (b).

Let $B^*(t\,|\,b)$ be the probability that an individual newly born at time 0 and known ultimately to suffer fate (b) has a life length less than or equal to t.

Let $D^*(t)$, $D^*(\infty)$, $D^*(t\,|\,d)$ be the corresponding probabilities for fate (d).

The probability that an individual born at time 0 suffers fate (b) in the interval $(t, t+dt)$ is

$$[1 - G(t)]\,\lambda(t)\,dt + o(dt).$$

Hence

$$B^*(t) = \int_0^t [1 - G(u)]\,\lambda(u)\,du,$$

and, since

$$B^*(t) = B^*(\infty)\,B^*(t\,|\,b),$$

$$B^*(t\,|\,b) = \int_0^t [1 - G(u)]\,\lambda(u)\,du \bigg/ \int_0^\infty [1 - G(u)]\,\lambda(u)\,du.$$

with a similar result for $D^*(t|d)$. It will be desirable in any attempt to fit the model to experimental data to give to $B^*(t|b)$ the form of the observed distribution of life lengths for those individuals whose life ultimately terminates in rate (b), and similarly for $D^*(t|d)$.

We can express $B^*(t)$ in terms of $B(t)$ and $D(t)$ as follows

Since
$$\lambda(t) = b(t)/[1-B(t)]$$

and
$$1-G(t) = [1-B(t)][1-D(t)],$$

therefore
$$B^*(t) = \int_0^t [1-D(u)]b(u)\,du,$$

and similarly
$$D^*(t) = \int_0^t [1-B(u)]d(u)\,du.$$

From the rigorous definition of $Z(t)$ (the population size) as a random variable $r_t(\omega)$ over a suitable sample space Ω, to which we have referred, we can show that

$$p_0(t) \equiv \Pr\{Z(t)=0\}$$

is a solution of the integral equation

$$p_0(t) = \int_0^t \{q_0(u) + q_2(u)[p_0(t-u)]^2\}g(u)\,du.$$

Clearly $p_0(t)$ must be a bounded solution because we must have $0 \leqslant p_0(t) \leqslant 1$, and, because $\Pr\{Z(t_2)=0\} \geqslant \Pr\{Z(t_1)=0\}$ for $t_2 > t_1$, $p_0(t)$ is a monotonic increasing function of t. Furthermore, it can be shown, by a method due to Bellman & Harris, that the bounded solution of the integral equation is unique.

Hence $p_0(t) \uparrow P \leqslant 1$ as $t \to \infty$.

Furthermore, $p_0(t)$ is the sum of an indefinite integral and the convolution of a bounded monotonic increasing function with a continuous function. Hence $p_0(t)$ is continuous.

Let $p_0(t) = P - \delta(t)$ so that $\delta(t) \downarrow 0$ as $t \to \infty$.

The integral equation can be written

$$P - \delta(t) = \int_0^t \{q_0(u) + P^2 q_2(u)\}g(u)\,du - 2P\int_0^t \delta(t-u)q_2(u)g(u)\,du + \int_0^t \delta^2(t-u)q_2(u)g(u)\,du.$$

Note that $\int_0^t q_2(u)g(u)\,du \leqslant 1$, that $\delta(t) \leqslant 1$, and that by Cauchy's principle of convergence $\int_s^t q_2(u)g(u)\,du < \epsilon$, where ϵ is any positive constant, for any $t > s$ when s is sufficiently large. Consider the second term in the equation above and divide the range of integration into two parts

$$\int_0^s \delta(t-u)q_2(u)g(u)\,du + \int_s^t \delta(t-u)q_2(u)g(u)\,du$$

$$\leqslant \int_0^s \delta(t-u)q_2(u)g(u)\,du + \int_s^t q_2(u)g(u)\,du \quad (\text{because } 0 \leqslant \delta(t) \leqslant 1),$$

$$\leqslant \int_0^s \delta(t-u)q_2(u)g(u)\,du + \epsilon \quad \text{for sufficiently large } s,$$

$$\leqslant \epsilon \int_0^s q_2(u)g(u)\,du + \epsilon \quad \text{for sufficiently large } t > s,$$

$$\leqslant 2\epsilon.$$

This argument and another similar one show that the last two terms on the right-hand side of the above integral equation tend to zero as $t \to \infty$, and so we get

$$P = \int_0^\infty \{q_0(u) + P^2 q_2(u)\} g(u)\, du.$$

Let

$$\rho = \int_0^\infty q_0(u) g(u)\, du \quad \text{and} \quad \sigma = \int_0^\infty q_2(u) g(u)\, du,$$

then, since by the assumption that $\int_0^t [\lambda(u) + \nu(u)]\, du \to \infty$ as $t \to \infty$, the function $G(t) \to 1$ as $t \to \infty$

$$P = \rho + \sigma P^2, \quad \text{where} \quad \rho + \sigma = 1.$$

The roots of this quadratic are $P = 1$, $P = \rho / \sigma$. If $\rho \geqslant \sigma$ we must therefore have $p_0(t) \to 1$ because the limit of $p_0(t)$ is not greater than 1.

On the other hand, if $\rho < \sigma$, the limit of $p_0(t)$ might be either ρ / σ or 1.

Suppose the limit of $p_0(t)$ is 1. Then, for some t_1, $0 < t_1 < \infty$, we must have

$$p_0(t_1) = \theta, \quad \text{where} \quad \rho / \sigma < \theta < 1.$$

Then, for this value of t,

$$\theta = \int_0^{t_1} \{q_0(u) + q_2(u)\, [p_0(t_1 - u)]^2\} g(u)\, du$$

$$\leqslant \int_0^{t_1} \{q_0(u) + \theta^2 q_2(u)\} g(u)\, du.$$

Now if the upper limit of integration on the right-hand side is increased, the right-hand side cannot decrease. Furthermore, if $q_0(t)\, g(t)$ or $q_2(t)\, g(t)$ is non-zero for $t > t_1$, it must *increase*, if we increase the range of integration to $(0, \infty)$.

In this case

$$\theta < \rho + \sigma \theta^2,$$

i.e.

$$\sigma \theta^2 - \theta + \rho > 0.$$

But θ lies between the roots ρ / σ and 1 of the left-hand side, which is a quadratic expression that must be negative inside the interval $(\rho / \sigma, 1)$. Hence we have a contradiction, and therefore $p_0(t)$ cannot take any value in the interval $(\rho / \sigma, 1)$.

On the other hand, if $q_0(t)\, g(t)$ and $q_2(t)\, g(t)$ vanish identically for $t > t_1$, we may have the equality

$$\theta = \rho + \sigma \theta^2,$$

but this implies $\theta = \rho / \sigma$ or $\theta = 1$, which again contradicts the assumption $\rho / \sigma < \theta < 1$.

Hence we see that $p_0(t) \uparrow \min \{\rho / \sigma, 1\}$ and this limit is the probability of extinction.

5. Example: the simplest Markovian birth and death process

To illustrate the methods of the preceding sections we shall obtain the integral equation, etc., for the birth and death process in which the probabilities are as follows:

For a particle alive at time t

probability to split into 2 new particles during $(t, t + dt) = \lambda\, dt + o(dt)$,

probability to die during $(t, t + dt) \qquad\qquad = \nu\, dt + o(dt)$,

where λ and ν are constants.

In this case $b(t) = \lambda e^{-\lambda t}$ and $d(t) = \nu e^{-\nu t}$. Hence we obtain

$$g(t) = (\lambda + \nu) e^{-(\lambda+\nu)t},$$

$$g_0(t) = \nu/(\lambda+\nu), \quad g_2(t) = \lambda/(\lambda+\nu),$$

and the integral equation is

$$\phi(z,t) = \int_0^t [\nu + \lambda\phi^2(z,t-u)] e^{-(\lambda+\nu)u} du + z e^{-(\lambda+\nu)t}.$$

For the observable life-length distributions we have

$$B^*(t\,|\,b) = \int_0^t \lambda e^{-(\lambda+\nu)u} du \Big/ \int_0^\infty \lambda e^{-(\lambda+\nu)u} du$$

$$= 1 - e^{-(\lambda+\nu)t}$$

$$= G(t),$$

and the same result holds for $D^*(t\,|\,d)$. For the extinction probabilities since

$$\sigma = B^*(\infty) = \lambda/(\lambda+\nu) \quad \text{and} \quad \rho = D^*(\infty) = \nu/(\lambda+\nu),$$

the probability of extinction is equal to 1 if $\lambda \leqslant \nu$ and is equal to ν/λ otherwise. These particular results are of course all well known.

6. CHOICE OF A PARTICULAR FORM FOR $b(t)$ AND $d(t)$

Although the functions that appear in the basic statement of the birth and death process are $\lambda(t)$ and $\nu(t)$, it is a clearer way of indicating the type of age dependence we are considering if we give first the frequency functions of the life-length distributions. In the ensuing sections of this paper we shall take

$$b(t) \equiv \frac{(k\lambda)^k}{(k-1)!} t^{k-1} e^{-k\lambda t},$$

$$d(t) \equiv \frac{(m\nu)^m}{(m-1)!} t^{m-1} e^{-m\nu t},$$

where λ and ν are positive constants.

Probability distributions of this type have been used before in population problems (see Kendall, 1952b) and also (much earlier) in the theory of queues (for an account of this see Erlang, 1948). A 'microscopic' interpretation of them has been given in the birth process, where only one fate is possible for each member of the population, i.e. in the binary fission process with no death effect. This interpretation is that at the birth of a new particle a process is started within it, which must go through a fixed number, k, of successive distinct phases. The times spent in the successive phases are independent random variables, each having distribution function $1 - e^{-k\lambda t}$. When the last phase is completed the particle subdivides and forms two new particles. It follows that the 'generation time' is distributed like $\chi^2_{2k}/(2k\lambda)$. Note that in giving this interpretation we do not intend to imply that if a real population is found to have a generation time with this distribution then a process of k phases is necessarily going on within the individuals. Our reason for adopting this form of distribution of generation time is that it appears to give a fairly good fit in certain populations (see, for example, Kendall (1948) and recent experimental results by Dr E. O. Powell (1955), and a real physical multiphase process may or may not be held to account for this.

We can extend this multi-phase interpretation to our model if we consider that at the birth of a new particle two such processes start within it, one being of k phases and the other of m phases. We suppose that the fate of the particle is decided as follows: if the k-phase process (which we might call 'maturity') is completed first, the particle undergoes fate (b) at the moment of completion, if the m-phase process (which we might call 'decay') is completed first, the particle undergoes fate (d).

An important property of the above distributions is that as the 'number of phases' k (or m) tends to infinity the distributions tend towards the corresponding deterministic distributions, while for all k (or m) they have the fixed mean value $1/\lambda$ (or $1/\nu$).

It follows from the above choice of $b(t)$ and $d(t)$ that

$$B(t) \equiv 1 - e^{-k\lambda t}\left[1 + k\lambda t + \ldots + \frac{(k\lambda t)^{k-1}}{(k-1)!}\right],$$

$$D(t) \equiv 1 - e^{-m\nu t}\left[1 + m\nu t + \ldots + \frac{(m\nu t)^{m-1}}{(m-1)!}\right],$$

$$\lambda(t) \equiv \frac{k\lambda}{(k-1)!}\frac{(k\lambda t)^{k-1}}{1 + k\lambda t + \ldots + \frac{(k\lambda t)^{k-1}}{(k-1)!}},$$

$$\nu(t) \equiv \frac{m\nu}{(m-1)!}\frac{(m\nu t)^{m-1}}{1 + m\nu t + \ldots + \frac{(m\nu t)^{m-1}}{(m-1)!}}.$$

Making use of properties of the above forms of $B(t)$ and $D(t)$ we shall express each integral equation in terms of convolutions of functions, and then reduce it to a linear differential equation with constant coefficients. The subsequent solution depends on solving the characteristic equation explicitly, and hence the equation for the mean population size.

In the next two sections we shall suppose that $k \geqslant 1$, but that $m = 1$. Roughly speaking this assumption means that births occur in an age-dependent manner but deaths occur at random. Because a complete solution is available and because the assumption may well be realistic in some applications, we give the explicit solution for the mean and derive the variance and coefficient of variation of the population size from it. We can compare our results (i) with Kendall's (1948) uniform multi-phase pure birth process by putting $\nu = 0$, and (ii) with Feller's Markovian birth and death process by putting $k = m = 1$. This leads, of course, to the simple form of this process in which the birth and death rates do not depend on the population size.

When the number of phases in both $B(t)$ and $D(t)$ exceeds 1 an explicit solution of the characteristic equation is not readily obtainable, so we give asymptotic formulae for the mean and variance of the population size in terms of the dominant root of the characteristic equation.

We shall find the following notation convenient.

Let
$$(\alpha) = \alpha e^{-\alpha t}.$$

Let
$$\int_0^t f(t-u)\alpha e^{-\alpha u}\,du = f(t) * (\alpha).$$

Thus $\alpha^2 t e^{-\alpha t} = (\alpha) * (\alpha)$. We shall write $(\alpha) * (\alpha) = (\alpha)^{*2}$, and, extending this analogy with powers,

$$\frac{\alpha^n}{(n-1)!}t^{n-1}e^{-\alpha t} = (\alpha) * (\alpha) * \ldots * (\alpha) = (\alpha)^{*n}.$$

Let

$$\int_0^t \frac{\alpha^n}{(n-1)!} u^{n-1} e^{-\alpha u} \, du = \int_0^t (\alpha)^{*n} \, du$$

$$= 1 \bullet (\alpha)^{*n}$$

be the distribution function of $\frac{1}{2}\chi_{2n}^2$, α, where 1 is the function that is constant and equal to unity for all t in $-\infty < t < \infty$.

Let $k\lambda + mv = \gamma$.

In this notation we can express the functions which occur in the integral equations as follows:

$$b(t)\,[1 - D(t)] = \frac{(k\lambda)^k}{(k-1)!} e^{-\gamma t} \left[t^{k-1} + mvt^k + \dots + \frac{(mv)^{m-1}}{(m-1)!} t^{m+k-2} \right]$$

$$= (k\lambda)^k \left[\frac{1}{\gamma^k} \frac{\gamma^k}{(k-1)!} t^{k-1} e^{-\gamma t} + mv \frac{k}{\gamma^{k+1}} \frac{\gamma^{k+1}}{k!} t^k e^{-\gamma t} + \dots \right.$$

$$\left. + \frac{(mv)^{m-1}}{(m-1)!} \frac{k(k+1)\dots(k+m-2)}{\gamma^{k+m-1}} \frac{\gamma^{k+m-1}}{(k+m-2)!} t^{k+m-2} e^{-\gamma t} \right]$$

$$= \left(\frac{k\lambda}{\gamma}\right)^k \left[(\gamma)^{*k} + \frac{mv}{\gamma} \frac{k}{1} (\gamma)^{*(k+1)} + \dots \right.$$

$$\left. + \left(\frac{mv}{\gamma}\right)^{m-1} \frac{k(k+1)\dots(k+m+2)}{(m-1)!} (\gamma)^{*(k+m-1)} \right].$$

Similarly

$$d(t)\,[1 - B(t)] = \left(\frac{mv}{\gamma}\right)^m \left[(\gamma)^{*m} + \frac{k\lambda}{\gamma} \frac{m}{1} (\gamma)^{*(m-1)} + \dots \right.$$

$$\left. + \left(\frac{k\lambda}{\gamma}\right)^{k-1} \frac{m(m+1)\dots(m+k-2)}{(k-1)!} (\gamma)^{*(m+k-1)} \right].$$

The observable life-length distributions corresponding to probabilities of the type we have adopted are given by

$$B^*(t) = \left(\frac{k\lambda}{\gamma}\right)^k \left[1 * (\gamma)^{*k} + \frac{mv}{\gamma} \frac{k}{1} 1 * (\gamma)^{*(k+1)} + \dots \right.$$

$$\left. + \left(\frac{mv}{\gamma}\right)^{m-1} \frac{k(k+1)\dots(k+m+2)}{(m-1)!} 1 * (\gamma)^{*(k+m-1)} \right],$$

whence $B^*(\infty)$ follows on noting that $1 * (\gamma)^{*(k+j)}$ tends to 1 as $t \to \infty$ for $j = 0, 1, \dots, m-1$. Hence we can see that the distribution function $B^*(t \mid b)$, and similarly $D^*(t \mid d)$, is the sum of constant multiples of several distribution functions each of χ^2 form, i.e. it is a mixture of χ^2 variables (see Robbins, 1948).

The probability of extinction is equal to 1 if

$$\left(\frac{k\lambda}{\gamma}\right)^k \left[1 + \frac{mv}{\gamma} \frac{k}{1} + \dots + \left(\frac{mv}{\gamma}\right)^{m-1} \frac{k(k+1)\dots(k+m-2)}{(m-1)!} \right] \leqslant \frac{1}{2}.$$

Note that when $k = m$ this is equivalent to $v \geqslant \lambda$.

The expression above consists of the first m terms of the negative binomial distribution whose generating function is $p^k(1-qs)^{-k}$ when we put $p \equiv k\lambda/\gamma$ and $q \equiv mv/\gamma$, and thus the given condition is equivalent to the statement that *the median of this negative binomial distribution shall be greater than or equal to m*.

7. The case of random death: mean population size

In this case, putting $m = 1$, we have

$$b(t)[1 - D(t)] = \left(\frac{k\lambda}{\gamma}\right)^k (\gamma)^{*k},$$

$$d(t)[1 - B(t)] = \frac{\nu}{\gamma}\left[(\gamma) + \frac{k\lambda}{\gamma}(\gamma)^{*2} + \ldots + \left(\frac{k\lambda}{\gamma}\right)^{k-1}(\gamma)^{*k}\right].$$

The corresponding observable life-length distributions are obtained as follows:

$$B^*(t) = \left(\frac{k\lambda}{\gamma}\right)^k 1 * (\gamma)^{*k}, \quad \text{whence} \quad B^*(\infty) = \left(\frac{k\lambda}{\gamma}\right)^k$$

and so

$$B^*(t \mid b) = 1 * (\gamma)^{*k},$$

which is of just the same form as $B(t)$ with the parameter $\gamma = k\lambda + \nu$ replacing $k\lambda$. The effect of deaths occurring at random, on the observable birth rate, is merely to alter the constant in the distribution without altering the form of the distribution whereby the dependence of the birth-rate on the age of the parent is expressed.

$$D^*(t) = \frac{\nu}{\gamma}\left[1 * (\gamma) + \frac{k\lambda}{\gamma} 1 * (\gamma)^{*2} + \ldots + \left(\frac{k\lambda}{\gamma}\right)^{k-1} 1 * (\gamma)^{*k}\right],$$

whence

$$D^*(\infty) = \frac{\nu}{\gamma}\left[1 + \frac{k\lambda}{\gamma} + \ldots + \left(\frac{k\lambda}{\gamma}\right)^{k-1}\right] = 1 - \left(\frac{k\lambda}{\gamma}\right)^k,$$

and so

$$D^*(t \mid d) = \frac{\nu}{\gamma}\left[1 - \left(\frac{k\lambda}{\gamma}\right)^k\right]^{-1}\left[1 * (\gamma) + \frac{k\lambda}{\gamma} 1 * (\gamma)^{*2} + \ldots + \left(\frac{k\lambda}{\gamma}\right)^{k-1} 1 * (\gamma)^{*k}\right].$$

Thus, as might be expected, although the actual death-rate is independent of the age of the individuals the effect of the age-dependent birth-rate is to produce an observable death-rate which *is* age-dependent.

The probability of extinction is equal to 1 if $\left(\frac{k\lambda}{\gamma}\right)^k \leqslant \frac{1}{2}$, i.e. if $\nu \geqslant k(2^{1/k} - 1)\lambda$, while if $\nu < k(2^{1/k} - 1)\lambda$, then

$$\text{Probability of extinction} = \left[1 - \left(\frac{k\lambda}{\gamma}\right)^k\right] \Big/ \left(\frac{k\lambda}{\gamma}\right)^k$$

$$= \left(1 + \frac{1}{k}\frac{\nu}{\lambda}\right)^k - 1.$$

In the case of random death the integral equation for the mean population size may be written in the 'convolution' notation as

$$\mu_1(t) = 2\left(\frac{k\lambda}{\gamma}\right)^k \mu_1(t) * (\gamma)^{*k} + 1 - \left(\frac{k\lambda}{\gamma}\right)^k 1 * (\gamma)^{*k} - \frac{\nu}{\gamma} 1 * \left[(\gamma) + \frac{k\lambda}{\gamma}(\gamma)^{*2} + \ldots + \left(\frac{k\lambda}{\gamma}\right)^{k-1}(\gamma)^{*k}\right],$$

where we have put $1 - G(t) = 1 - 1 * g(t)$.

Let us define a differential operator,

$$\delta \equiv 1 + \frac{1}{\gamma}\frac{d}{dt},$$

so that

$$\delta[f(t) * (\gamma)] \equiv f(t).$$

Put $t = 0$ in the integral equation and in the relations resulting from applying the operators $\delta, \delta^2, \ldots, \delta^{k-1}$ to it. This gives

$$\mu_1(0) = 1,$$

$$\delta\mu_1(0) = \frac{k\lambda}{\gamma},$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$\delta^{k-1}\mu_1(0) = \left(\frac{k\lambda}{\gamma}\right)^{k-1}.$$

whence it follows that

$$\left(\frac{d}{dt}\right)^j \mu_1(t)\bigg|_{t=0} = (-\nu)^j \quad (j = 0, 1, \ldots, k-1).$$

While δ^k gives the differential equation of the mean

$$\delta^k\mu_1(t) = 2\left(\frac{k\lambda}{\gamma}\right)^k \mu_1(t) + 1 - \left(\frac{k\lambda}{\gamma}\right)^k + \frac{\nu}{\gamma}\left[1 + \frac{k\lambda}{\gamma} + \ldots + \left(\frac{k\lambda}{\gamma}\right)^{k-1}\right].$$

The constant terms sum to zero so we get

$$\left\{\left(1 + \frac{1}{\gamma}\frac{d}{dt}\right)^k - 2\left(\frac{k\lambda}{\gamma}\right)^k\right\}\mu_1(t) = 0.$$

The characteristic equation for the above differential equation is

$$\left(1 + \frac{x}{\gamma}\right)^k - 2\left(\frac{k\lambda}{\gamma}\right)^k = 0.$$

If ω is the primitive kth root of unity this gives

$$x = 2^{1/k}k\lambda\omega^j - \gamma, \quad \text{where} \quad j = 0, 1, \ldots, k-1.$$

These roots lie on a circle in the complex plane, centre $-\gamma$, radius $2^{1/k}k\lambda$. The root of largest real part is

$$x = 2^{1/k}k\lambda - \gamma = k(2^{1/k}-1)\lambda - \nu.$$

To obtain an explicit solution for the mean population size it is convenient to consider the mean population size which results when there is no death effect ($\nu = 0$). In this case the system reduces to the uniform multi-phase birth process described by Kendall (1948, 1952b). If we call the mean population size $\mu_1^*(t)$ when $\nu = 0$ we see that it satisfies the equation

$$\left\{\left(1 + \frac{1}{k\lambda}\frac{d}{dt}\right)^k - 2\right\}\mu_1^*(t) = 0,$$

together with the end conditions

$$\mu_1^*(0) = 1 \quad \text{and} \quad \left(\frac{d}{dt}\right)^j \mu_1^*(0) = 0 \quad \text{for} \quad j = 1, 2, \ldots, k-1.$$

The solution of this system is

$$\mu_1^*(t) = \frac{1}{2k}\sum_{j=0}^{k-1}\frac{2^{1/k}\omega^j}{2^{1/k}\omega^j - 1}\exp\left[k(2^{1/k}\omega^j - 1)\lambda t\right]$$

$$= e^{-k\lambda t}\sum_{r=0}^{\infty}2^r\left[\frac{(k\lambda t)^{rk}}{(rk)!} + \frac{(k\lambda t)^{rk+1}}{(rk+1)!} + \ldots + \frac{(k\lambda t)^{rk+k-1}}{(rk+k-1)!}\right].$$

It follows that if we put
$$\mu_1(t) = e^{-\nu t}\mu_1^*(t),$$

then $\mu_1(t)$ will satisfy the differential equation, together with the end-conditions, that we have derived for it above.

The value of the mean population size for large t depends on whether the dominant root of the characteristic equation, i.e. the real root $(2^{1/k}-1)k\lambda-\nu$ is less than, equal to, or greater than 0. We shall consider these cases in turn.

(i) $\nu > k(2^{1/k}-1)\lambda$. The mean population size tends to zero like $e^{-\alpha t}$, where α is a positive constant, and it follows from this or from the value of $B^*(\infty)$ that the probability of extinction is 1.

(ii) $\nu = k(2^{1/k}-1)\lambda$. The probability of extinction is again equal to 1 but we obtain a finite limiting mean population size
$$\mu_1(t) \to C_k, \quad \text{where} \quad C_k = \frac{2^{1/k}}{2k(2^{1/k}-1)}.$$

(iii) $\nu < k(2^{1/k}-1)\lambda$. The probability of extinction is equal to $\left(1+\frac{1}{k}\frac{\nu}{\lambda}\right)^k - 1 < 1$. The mean population size will increase exponentially and we can write the following asymptotic expression for it:
$$\mu_1(t) \sim \frac{2^{1/k}}{2k(2^{1/k}-1)} \exp\left[\{(2^{1/k}-1)k\lambda-\nu\}t\right].$$

8. THE CASE OF RANDOM DEATH: COEFFICIENT OF VARIATION OF THE POPULATION SIZE

Using the convolution notation we may express the integral equation connecting the mean with the second factorial moment of the population size as follows:
$$\mu_2(t) = 2\left(\frac{k\lambda}{\gamma}\right)^k \{\mu_1(t)\}^2 * (\gamma)^{*k} + 2\left(\frac{k\lambda}{\gamma}\right)^k \mu_2(t) * (\gamma)^{*k}.$$

By applying the differential operator δ as in the case of the equation for the mean this reduces to
$$\left\{\left(1+\frac{1}{\gamma}\frac{d}{dt}\right)^k - 2\left(\frac{k\lambda}{\gamma}\right)^k\right\}\mu_2(t) = 2\left(\frac{k\lambda}{\gamma}\right)^k \{\mu_1(t)\}^2,$$

i.e. multiplying through by γ^k and factorizing the operator
$$\prod_{r=0}^{k-1}\left\{\frac{d}{dt} - (2^{1/k}\omega^r - 1)k\lambda + \nu\right\}\mu_2(t) = 2(k\lambda)^k\{\mu_1(t)\}^2.$$

As with the mean, the behaviour of the second factorial moment as $t \to \infty$ depends on whether the dominant root of the characteristic equation is negative, zero, or positive, and we shall examine these cases in turn.

(i) $\nu > k(2^{1/k}-1)\lambda$. All terms in the solution for the second factorial moment will tend to zero faster than $e^{-\alpha t}$, where α is some positive constant.

(ii) $\nu = k(2^{1/k}-1)\lambda$. The term $\{\mu_1(t)\}^2$ on the right-hand side of the equation tends to the finite limit C_k^2 as $t \to \infty$, while the equation is
$$\frac{d}{dt}\prod_{r=1}^{k-1}\left\{\frac{d}{dt} - k(2^{1/k}\omega^r - 1)\lambda + \nu\right\}\mu_2(t) = 2(k\lambda)^k\{\mu_1(t)\}^2.$$

Hence for large t we shall have an asymptotic solution.

$$\mu_2(t) \sim \frac{2(k\lambda)^k}{\prod\limits_{r=1}^{k-1}\{v - k(2^{1/k}\omega^r - 1)\lambda\}} e^{\frac{2}{k}t}$$

$$= \frac{2(k\lambda)^k}{k(k\lambda + v)^{k-1}} C_k^2 t.$$

Hence the variance tends to infinity, but does so to the order of t rather than t^2 and in this the behaviour of this age-dependent system resembles that of the simple Markovian birth and death process when the parameters λ and r in it are equal.

Note that as $k \to \infty$ we have the condition $r = \lambda \log_e 2 \approx 0.69315\lambda$ for a finite limiting mean population size.

(iii) $v < k(2^{1/k} - 1)\lambda$. The dominant term in the complementary function will be of the order of $\exp[\{k(2^{1/k} - 1)\lambda - r\}t]$, but the particular integral will contain a term in $\exp[\{k(2^{1/k} - 1)\lambda - v\}2t]$ which will therefore give the asymptotic value of the complete solution. Formally we can write the particular integral as

$$\mu_2(t) = \frac{2(k\lambda)^k}{\left(\gamma + \dfrac{d}{dt}\right)^k - 2(k\lambda)^k}\{\mu_1(t)\}^2,$$

and hence

$$\mu_2(t) \sim \frac{2(k\lambda)^2}{[k(2^{(k-1)/k} - 1)\lambda - r]^k - 2(k\lambda)^k}\{\mu_1(t)\}^2.$$

Let $V_n(t)$ be the coefficient of variation of the population size. Then

$$V_n(t) = \frac{\mu_2(t)}{\{\mu_1(t)\}^2} + \frac{1}{\mu_1(t)} - 1,$$

and if we introduce the asymptotic values for the mean and second factorial moment, we obtain for the coefficient of variation for large t

$$V_n^2 = \lim_{t\to\infty} V_n^2(t) = \frac{2}{\left(2^{(k+1)/k} - 1 - \dfrac{v}{k\lambda}\right)^k - 2} - 1$$

$$= \frac{2}{\left(1 + \dfrac{2\alpha_k - v/k}{k}\right)^k - 2} - 1,$$

where $\alpha_k \equiv k(2^{1/k} - 1)$.

When $k \to \infty$, the process tends to resemble one in which deaths occur at random but fission occurs deterministically when the parent individual attains a certain fixed age equal to $1/\lambda$. The limit as $k \to \infty$ of the above coefficient of variation, which is itself a limit for large t, is thus of interest. Noting that $\alpha_k \to \log_e 2$ and that $\frac{1}{2} < e^{-v\lambda} \leqslant 1$, with equality only when $v = 0$ we obtain the limit

$$\lim_{k\to\infty}(V_n^2) = 2\left(\frac{1 - e^{-v/\lambda}}{2e^{-v/\lambda} - 1}\right).$$

If V_r is the coefficient of variation of the distribution $B(t)$, then $V_r = 1/\sqrt{k}$ and so $V_n/V_r = V_n\sqrt{k}$ will have a finite limit only if $V_n \to 0$, as $k \to \infty$, i.e. if $v = 0$. This point will be of significance if any attempt is made to estimate V_r from observations of V_n, the latter being easier to observe, for instance, in bacterial populations.

9. Birth- and Death-Rates both Age-Dependent

We have seen that when $k > 1$ and $m > 1$ the instantaneous probabilities of death and of fission both depend on the age of the individual considered. We will first investigate the behaviour of the mean population size under these conditions. The integral equation for the mean population size is

$$\mu_1(t) = 2\left(\frac{k\lambda}{\gamma}\right)^k \mu_1(t) * \left[(\gamma)^{*k} + \frac{m\nu}{\gamma}\frac{k}{1}(\gamma)^{*(k+1)} + \dots \right.$$

$$+ \left(\frac{m\nu}{\gamma}\right)^{m-1}\frac{k(k+1)\dots(k+m-2)}{(m-1)!}(\gamma)^{*(k-m-1)}\Bigg]$$

$$+ [1 - B(t)][1 - D(t)].$$

Now
$$[1 - B(t)][1 - D(t)] \equiv e^{-\gamma t}P(t),$$

where $P(t)$ is a polynomial of degree $k + m - 2$ in t. Hence

$$\delta^{k+m-1}[1 - B(t)][1 - D(t)] \equiv \left(1 + \frac{1}{\gamma}\frac{d}{dt}\right)^{k+m-1}e^{-\gamma t}P(t) \equiv 0.$$

Thus, if we apply the operator δ^{k+m-1} to the above integral equation we obtain the following homogeneous differential equation for the mean population size

$$F\left(\frac{d}{dt}\right)\mu_1(t) \equiv \left\{\delta^{k+m-1} - 2\left(\frac{k\lambda}{\gamma}\right)^k\left[\delta^{m-1} + \frac{m\nu}{\gamma}\frac{k}{1}\delta^{m-2} + \dots \right.\right.$$

$$\left.\left. + \left(\frac{m\nu}{\gamma}\right)^{m-1}\frac{k(k+1)\dots(k+m-2)}{(m-1)!}\right]\right\}\mu_1(t) = 0.$$

The characteristic equation is $F(x) = 0$ and if we put $y = 1 + x/\gamma$ this is

$$H(y) \equiv y^{k+m-1} - 2\left(\frac{k\lambda}{\gamma}\right)^k\left[y^{m-1} + \frac{m\nu}{\gamma}\frac{k}{1}y^{m-2} + \dots + \left(\frac{m\nu}{\gamma}\right)^{m-1}\frac{k(k+1)\dots(k+m-2)}{(m-1)!}\right] = 0.$$

It does not seem as easy in this case as it was in the case of random death ($m = 1$) to obtain an explicit solution of the characteristic equation. We will obtain an asymptotic solution for the mean population size, and will consider only the case when the probability of extinction is less than 1. The condition for the probability of extinction to be less than 1 is $\rho < \sigma$ and this is the same as $1 - 2B^*(\infty) < 0$, i.e. $H(1) < 0$. Since $H(y) \to \infty$ as $y \to \infty$ it follows that the characteristic equation $H(y) = 0$ must have at least one real root greater than 1. There is just one change of sign in $H(y)$ so by Descartes' rule of signs it follows that $H(y) = 0$ has a unique real positive root, say w, which is simple and greater than 1.

We can show that w is larger than the real part of any other root of $H(y) = 0$. Let $z = u + iv$ be any complex root. We can ignore real roots and complex roots with $u \leqslant 0$. We have, rearranging $H(z) = 0$

$$2\left(\frac{k\lambda}{\gamma}\right)^k\left\{\frac{1}{z^k} + \frac{m\nu}{\gamma}\frac{k}{1}\frac{1}{z^{k+1}} + \dots + \left(\frac{m\nu}{\gamma}\right)^{m-1}\frac{k(k+1)\dots(k+m-2)}{(m-1)!}\frac{1}{z^{k+m-1}}\right\} = 1,$$

whence, taking moduli,

$$2\left(\frac{k\lambda}{\gamma}\right)^k\left\{\frac{1}{|z|^k} + \frac{m\nu}{\gamma}\frac{k}{1}\frac{1}{|z|^{k+1}} + \dots + \left(\frac{m\nu}{\gamma}\right)^{m-1}\frac{k(k+1)\dots(k+m-2)}{(m-1)!}\frac{1}{|z|^{k+m-1}}\right\} \geqslant 1.$$

Now if z is complex, $z \neq 0$, and so $|z| > u$. The left hand side, considered as a polynomial in $|z|^{-1}$ is strictly increasing for any argument greater than 0, because all coefficients are positive, so if we substitute u^{-1} for $|z|^{-1}$ we can drop the sign of equality. Thus

$$2\left(\frac{k\lambda}{\gamma}\right)^{k}\left\{\frac{1}{u^{k}}+\frac{m\nu}{\gamma}\frac{k}{1}\frac{1}{u^{k+1}}+\cdots+\left(\frac{m\nu}{\gamma}\right)^{m-1}\frac{k(k+1)\ldots(k+m-2)}{(m-1)!}\frac{1}{u^{k+m-1}}\right\}$$

$$>1+2\left(\frac{k\lambda}{\gamma}\right)^{k}\left\{\frac{1}{u^{k}}+\frac{m\nu}{\gamma}\frac{k}{1}\frac{1}{u^{k+1}}+\cdots+\left(\frac{m\nu}{\gamma}\right)^{m-1}\frac{k(k+1)\ldots(k+m-2)}{(m-1)!}\frac{1}{u^{k+m-1}}\right\}.$$

whence $w > u$.

It follows that $F(x) = 0$ has a root r which is unique, simple, real and positive and larger than the real part of any other root of $F(x) = 0$. Hence we have, as the asymptotic expression for the mean population size

$$\mu_1(t) \sim ce^{rt},$$

where c is a constant which could in theory be evaluated, though to do so we should require the other roots of the characteristic equation, and the appropriate end-conditions. However, it is not essential to evaluate this constant, because it will disappear from the limiting form of the coefficient of variation of the population size.

To write down the differential equation for the second factorial moment of the population size we note that

$$\left(1+\frac{x}{\gamma}\right)^{k+m-1}-F(x) \equiv 2\left(\frac{k\lambda}{\gamma}\right)^{k}\left\{\left(1+\frac{x}{\gamma}\right)^{m-1}+\frac{m\nu}{\gamma}\frac{k}{1}\left(1+\frac{x}{\gamma}\right)^{m-2}\right.$$

$$\left.+\ldots+\left(\frac{m\nu}{\gamma}\right)^{m-1}\frac{k(k+1)\ldots(k+m-2)}{(m-1)!}\right\}.$$

The integral equation for the second factorial moment is similar to the one which arose in the case of random death, but it contains more terms.

It can be reduced to a differential equation in just the same way, by applying the operator $\delta \equiv \left(1+\frac{1}{\gamma}\frac{d}{dt}\right)$, and the result is

$$F\left(\frac{d}{dt}\right)\mu_2(t) = \left\{\left(1+\frac{1}{\gamma}\frac{d}{dt}\right)^{k+m-1}-F\left(\frac{d}{dt}\right)\right\}\{\mu_1(t)\}^2.$$

We can see from the solution of the equation for the mean that the complementary function for this equation will have a dominant term of the order of e^{rt}, where r is the same dominant root of the characteristic equation. However, owing to the appearance of $\{\mu_1(t)\}^2$ on the right-hand side we see that the particular integral will contain a term of the order of e^{2rt} which will be the dominant term in the complete solution. Evaluating this term we obtain

$$\mu_2(t) \sim \frac{(1+2r/\gamma)^{k+m-1}-F(2r)}{F(2r)}c^2 e^{2rt}.$$

Note that the denominator $F(2r) \neq 0$.

It follows that the coefficient of variation of the population size for large t is given by

$$V_n^2 = \lim_{t\to\infty} V_n^2(t) = \frac{(1+2r/\gamma)^{k+m-1}}{F(2r)} - 2.$$

In conclusion I should like to thank Mr D. G. Kendall for his continued help and encouragement throughout the preparation of this paper.

REFERENCES

BELLMAN, R. & HARRIS, T. E. (1948). On the theory of age-dependent stochastic branching processes. *Proc. Nat. Acad. Sci., Wash.*, **34**, 601–4.

BELLMAN, R. & HARRIS, T. E. (1952). On age-dependent binary branching processes. *Ann. Math.* **55**, 280–95.

ERLANG, A. K. (1948) *The Life and Works of A. K. Erlang.* Edited by E. Brockmeyer, H. L. Halstrøm and Arne Jensen. Copenhagen: *Trans. Danish Acad. Tech. Sciences*, 1948, No. 2.

FELLER, W. (1939). Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung. *Acta biotheor., Leiden*, **5**, 11–40.

FELLER, W. (1950). *An Introduction to Probability Theory and its Applications.* New York: John Wiley and Sons, Inc.

HARRIS, T. E. (1951). Some mathematical models for branching processes. *Proc. 2nd Berkeley Symposium on Mathematical Statistics and Probability*, pp. 305–28.

KENDALL, D. G. (1948). On the role of variable generation time in the development of a stochastic birth process. *Biometrika*, **35**, 316–30.

KENDALL, D. G. (1952a). On the choice of a mathematical model to represent normal bacterial growth. *J. R. Statist. Soc.* B, **14**, 41–4.

KENDALL, D. G. (1952b). Les processus stochastiques de croissance en biologie. *Ann. Inst. Poincaré.* **13**, 43–108.

POWELL, E. O. (1955). Some features of the generation times of individual bacteria. *Biometrika*, **42**, 16–44.

RAMAKRISHNAN, ALLADI (1951). Some simple stochastic processes. *J. R. Statist. Soc.* B, **13**, 131–40.

REID, A. T. (1953). An age-dependent stochastic model of population growth. *Bull. Math. Biophys.* **15**, 361–5.

ROBBINS, H. (1948). Mixture of distributions. *Ann. Math. Statist.* **19**, 360–69.

A LARGE-SAMPLE BIOASSAY DESIGN WITH RANDOM DOSES AND UNCERTAIN CONCENTRATION*

By FRED C. ANDREWS and HERMAN CHERNOFF

The University of Nebraska and Stanford University

1. The Problem

The problem discussed here is that of designing an experiment for the purpose of estimating a parameter in a dose-response curve when the doses administered cannot be known with exactness.

To introduce the problem let us consider the following extreme example. A biologist has developed a strain of bacteria. He believes that this strain is so virulent that a dose of one organism applied to a test animal will lead to a response with a probability of the order of magnitude of $0 \cdot 2$. He wishes to estimate the virulence of this strain of bacteria more precisely. For experimentation he has available thirty test animals and 10 ml. of material containing this strain of bacteria in suspension. The concentration of bacterial organisms in this material is about four organisms per millilitre. but this latter number might be unreliable. The immediate problem confronting the biologist is one of design. Since he is not sure of the concentration he must use a portion of his material for a plate count (Clifton, 1950, pp. 243–4) to estimate the concentration. Then he must allocate portions of his remaining material among the test animals to determine virulence. The amount of material to be used in the plate count cannot be too large. for then the remainder will not be an adequate amount of dose material. Some balancing allocation between these two parts of the experiment must be reached or the entire experiment will not yield its maximum amount of information.

For the sake of the reader who prefers to peek at the last page of a mystery story. the optimal design corresponding to the biologist's problem (when completely formulated) is the following. To determine concentration $3 \cdot 1$ ml. of material are used. The remaining $6 \cdot 9$ ml. are divided equally among the thirty test animals.

According to this design about twenty-eight organisms are divided among thirty animals. It is obvious that not all of the animals are assured of receiving organisms. In fact, the dosage is random, and some animals might receive none. some might receive one and still others might receive two or more organisms. This points out that in our problem exact doses are not known for two reasons. First, the concentration is at best an estimate based on the limited amount of material assayed. Secondly. even if the concentration was known, the exact dose administered to an animal is a random variable which the biologist cannot observe directly. This second situation will be referred to as dosage subject to error. Dosage subject to error occurs frequently in practice but is seldom treated in theory (see Haley, 1953), although such theory is especially relevant when the doses are small.

The problem facing the biologist can be formulated briefly by stating that the fractions $f_0, f_1, f_2, \ldots, f_{30}$ must be determined with the understanding that f_0 is the fraction of the

material to be used in the plate count and f_i the part to be used to dose the ith test animal. Of course these fractions are to be determined in such a manner that an 'optimal' design for estimating the dose response curve will result.

2. Model

In order to formulate completely this design problem it is necessary to specify carefully the nature of the probability distributions involved.

First, the dose-response curve must be specified. In this paper, we shall treat only the case where our biologist has reason to believe that if d organisms are administered to a test animal the probability of no response is given by

$$p_d = (1-\alpha)^d. \tag{1}$$

The parameter α represents the probability that a dose of one organism will lead to a positive response. This dose-response curve is said to be exponential because the dose d appears in the exponent. This formula may also be written

$$p_d = e^{d \log_e (1-\alpha)}. \tag{2}$$

This model has previously been discussed in the literature by Goldberg & Watkins (1952), Druett (1952) and Peto (1953), and in a different context by Cochran (1950).

Secondly, we must consider the distribution of the number of organisms appearing in a certain amount of the material. Frequently the law of small numbers is assumed in this situation (see Worcester, 1954) yielding: *If the available material is taken from a source where the concentration is c organisms per millilitre, the number of organisms x appearing in a sample of r millimetres is a Poisson random variable with mean cr. That is,*

$$P(x=i) = e^{-cr}\frac{(cr)^i}{i!} \quad (i = 0, 1, 2, ...). \tag{3}$$

Furthermore, if several samples are taken from the source (without replacement) the number of organisms in these samples are independently distributed.

As an example, suppose that one-half of the tube of 10 ml. were used in the plate count. Suppose also that the concentration of the source was exactly 4 organisms per millilitre. Then the number of organisms in the part plated would be a Poisson random variable with mean (expectation) $\frac{1}{2} . 10 . 4 = 20$. When the dose d is a Poisson random variable with expectation D, we shall say that D is the nominal dose. Then

$$P(d=i) = \frac{e^{-D} D^i}{i!} \quad (i = 0, 1, 2, ...). \tag{4}$$

We might repeat here that in our problem the biologist does not know the concentration and must use the plate count to estimate the concentration of the source or equivalently the expected number of organisms in the test tube. Let λ be a notation for the expected number of organisms in the test-tube. If then the biologist uses a fraction f of his material as a dose, the nominal dose is $f\lambda$.

3. The effect of error in dose

As was pointed out in the introduction, the exact number of organisms in a dose is a random variable. Ordinarily, as in the case with the probit or logit models, the randomness of the dose very seriously complicates the mathematical situation, as is quite evident from reading

Haley (1953). It is, however, fortunate that with the use of the exponential model this complication is minimized.

Suppose that a nominal dose D is applied, then the probability of a negative response is given by

$$P_D = \sum_{i=0}^{\infty} P(d=i)(1-\alpha)^i = e^{-aD}.$$

Contrast this with the probability of a negative response when the exact dose is d, namely,

$$p_d = (1-\alpha)^d = e^{\log_e(1-\alpha)d}.$$

It is fundamental here to see that the response curve for a nominal dose is also exponential in form. The only difference in the two curves is that $\log_e(1-\alpha)$ in the exact dose curve is replaced by $-\alpha$ in the nominal dose curve. (For small α, $\log_e(1-\alpha)$ is almost equal to $-\alpha$.)

4. THE NATURE OF AN ASYMPTOTICALLY LOCALLY OPTIMAL DESIGN. AN EXAMPLE

The problem we have posed in § 1 yields a solution which might be termed a large-sample locally optimal design. To illustrate what we mean by these terms, we shall use a simpler problem which has already been treated.

Suppose that it is desired to dose a large number of test animals with organisms from a material with *known concentration* and that there is an unlimited supply of organisms available. This problem reduces then to selecting nominal doses $D_1, D_2, D_3, \ldots$ to administer to the animals. Once such nominal doses are selected and observations obtained, one may apply maximum-likelihood or some other asymptotically efficient estimation technique. It is known (though not trivial) how to apply the maximum-likelihood method of estimation and how to obtain the asymptotic variance of the resulting estimate of α. By an asymptotically optimal design we mean one which minimizes the asymptotic variance. It turns out for the exponential model that the asymptotic variance is minimized when all the doses are the same and these are equal to

$$D = \frac{1 \cdot 6}{\alpha}.$$

Here the local character of the problem becomes evident. That is, the asymptotically optimal design to estimate α depends on α. But it is clear that if α were known, there would be no point to the experiment. This paradoxical situation is resolved by the following consideration. In practice, the biologist will have a rough idea of the order of magnitude of α. Suppose he felt he could guess α of the order of magnitude of $0 \cdot 2$, then he could use the dose $1 \cdot 6/0 \cdot 2$ in place of $1 \cdot 6/\alpha$. It can easily be shown that the use of the wrong dose level increases the asymptotic variance of the estimate of α. However, this increase is very gradual. If we measure the dose in terms of percentage of the asymptotically optimal dose and the corresponding variance similarly we obtain the following table:

Dose/optimal dose	Variance/optimal variance
0·50	1·25
0·75	1·02
1·00	1·00
1·25	1·02
1·50	1·12

It is clear that if the biologist has any reasonable educated guess about α he could use the design which would be optimal if his guess were the correct value of α. If, however, his best guess might be off by a factor of three or more, alternative procedures would have to be considered. For example, he might use a small preliminary experiment involving several dose levels which are very highly spread out. Using the estimate of α based on this preliminary experiment he could then apply the locally optimal design.

The gradual increase in variance divided by optimal variance shows not only the applicability of the locally optimal design. It can also be used to show that designs where dose optimal dose takes on several values from 0·75 to 1·25 are almost optimal. While such designs lead to more complexity in the calculation of maximum-likelihood estimates they are useful to biologists who have reservations about the applicability of the exponential model to their problem.

5. Results

As may have been expected, the large-sample optimal solution to our design problem involving an assay is a local solution. In fact, it will be shown in the appendix that the solution has the following characteristics:

(1) Every test animal is given an equal fraction of the test-tube. According to the dictates of the mathematical model there is no advantage in limiting the number of animals used to allow larger individual doses. Of course costs of obtaining individual animals would forestall the use of an unlimited number of animals in any one experiment.

(2) The fraction of the test-tube administered to test animals is given by β^0 which is a function of α (probability that one organism will cause a positive reaction), λ (the expected number of organisms in the test tube), and s (the number of test animals available).

This functional dependence is rather complicated, but β^0 is very well approximated by β^* which is obtained from the following simple formula:

$$\beta^0 \approx \beta^* = \text{minimum}\left(\frac{1}{1+\sqrt{\alpha}}, \frac{1\cdot 6s}{\alpha\lambda}\right). \tag{5}$$

Hence if the numbers given in the statement were relevant we would have

$$\lambda \approx 40, \quad \alpha \approx 0\cdot 2, \quad s = 30, \quad \frac{1}{1+\sqrt{\alpha}} \approx 0\cdot 69, \quad \frac{1\cdot 6s}{\alpha\lambda} \approx 6, \quad \beta^* \approx 0\cdot 69.$$

A remark about how β^* was obtained might be pertinent here. If a very large number of organisms were available it would be desirable to give each animal a dose of about $1\cdot 6/\alpha$ and to use the large amount remaining for the plate count. Then the fraction of the test-tube used for doses would be about $1\cdot 6s/(\alpha\lambda)$. However, if there are very few organisms available, it turns out that $\sqrt{\alpha}/(1+\sqrt{\alpha})$ is approximately the right fraction to use in plate counting, leaving the fraction $1/(1+\sqrt{\alpha})$ for doses.

(3) The local nature of the solution is not a serious handicap to applications.

As in the example treated in §3, the asymptotic variance undergoes a relatively small change even when the fraction devoted to doses is changed by a considerable amount. This can be seen by referring to Figs. 1 and 2 and Table 1. Table 1 gives β^0 as a function of α and λ for $s = 30$. In addition, there is presented β, $\bar\beta$ and $\sigma^2(\dot{x})$, where $\beta < \beta^0 < \bar\beta$.

β and $\bar\beta$ are fractions devoted to doses which yield designs which are 80% efficient (efficiency measured in terms of asymptotic variance). $\sigma^2(\dot{x})$ is the asymptotic variance using β^0. The relatively large spread between β and $\bar\beta$ is an indication of the wide applicability of the local solution.

In Fig. 1, β, β^0, β^* and $\bar{\beta}$ are drawn as functions of λ for $\alpha = 0\cdot10$ and $s = 30$. Similarly, in Fig. 2, β, β^0, β^* and $\bar{\beta}$ are obtained for $\alpha = 0\cdot04$. These figures indicate the goodness of the approximation of β^* to β^0.
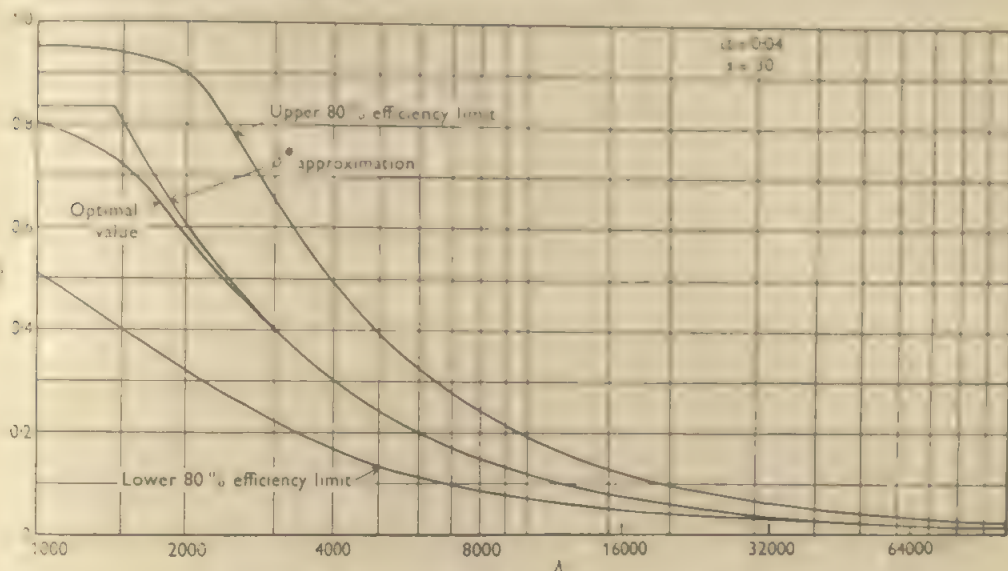
Fig. 1. Comparison of β values as a function of λ.
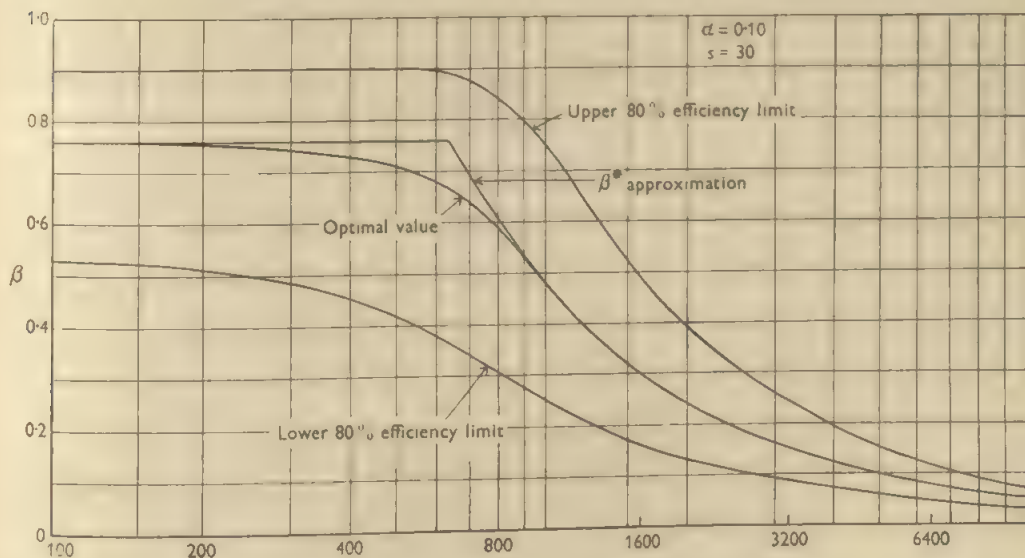
Fig. 2. Comparison of β values as a function of λ.

(4) The solution provides a very simple computational method of estimating α.

If all the test animals are given the same dose D, the proportion $\hat{p}$ of test animals which do not react is the maximum-likelihood estimate of $e^{-\alpha D}$. Hence, the estimate of α is given by

$$\hat{\alpha} = -\frac{1}{D}\log_e \hat{p} = \frac{-2\cdot30}{D}\log_{10}\hat{p}.$$

Table 1. *Various β values and the minimum asymptotic variance of α̂ for the case of s = 30 test animals*

λ \ α	0·001	0·005	0·020	0·04	0·06	0·08	0·10	0·15	0·20	λ
	*β β° β̄ σ²(α̂)									
100					0·58 0·80 0·92 10^{-3}(0·993332)	0·55 0·78 0·91 10^{-2}(0·143101)	0·53 0·76 0·90 10^{-2}(0·191414)	0·48 0·72 0·87 10^{-2}(0·331126)	0·45 0·69 0·85 10^{-2}(0·497069)	100
250					0·56 0·80 0·92 10^{-3}(0·440869)	0·53 0·77 0·91 10^{-3}(0·654412)	0·50 0·75 0·90 10^{-3}(0·900903)	0·44 0·70 0·88 10^{-2}(0·167033)	0·38 0·65 0·85 10^{-2}(0·267765)	250
500				0·58 0·83 0·94 10^{-3}(0·147557)	0·52 0·79 0·93 10^{-3}(0·265671)	0·47 0·75 0·91 10^{-3}(0·417767)	0·42 0·71 0·90 10^{-3}(0·608201)	0·32 0·57 0·85 10^{-2}(0·127089)	0·25 0·45 0·73 10^{-2}(0·220808)	500
1,000		0·73 0·93 0·96 10^{-3}(0·6617·25)	0·62 0·87 0·96 10^{-4}(0·342318)	0·51 0·80 0·95 10^{-4}(0·973386)	0·40 0·70 0·92 10^{-3}(0·199053)	0·31 0·57 0·87 10^{-3}(0·344729)	0·25 0·47 0·75 10^{-3}(0·533669)	0·17 0·32 0·51 10^{-2}(0·119120)	0·13 0·24 0·39 10^{-2}(0·211150)	1,000
1,500	0·70 0·97 — 10^{-6}(0·7284)	0·73 0·93 0·98 10^{-3}(0·42732)	0·59 0·86 0·96 10^{-4}(0·264795)	0·40 0·72 0·04 10^{-4}(0·866463)	0·28 0·52 0·83 10^{-3}(0·190346)	0·21 0·39 0·64 10^{-3}(0·338499)	0·17 0·32 0·52 10^{-3}(0·524522)	0·12 0·21 0·34 10^{-2}(0·117719)	0·09 0·16 0·26 10^{-2}(0·209062)	1,500
2,000	0·70 0·97 — 10^{-6}(0·5492)	0·72 0·93 0·95 10^{-5}(0·333001)	0·54 0·85 0·97 10^{-4}(0·231948)	0·32 0·58 0·90 10^{-4}(0·843020)	0·22 0·39 0·64 10^{-3}(0·188297)	0·16 0·30 0·48 10^{-3}(0·333991)	0·13 0·24 0·39 10^{-3}(0·521297)	0·09 0·16 0·26 10^{-2}(0·117150)	0·07 0·12 0·19 10^{-2}(0·208160)	2,000
2,500	0·79 0·97 — 10^{-6}(0·4428)	0·71 0·93 0·98 10^{-5}(0·27695)	0·48 0·81 0·97 10^{-4}(0·217409)	0·26 0·47 0·77 10^{-4}(0·835757)	0·17 0·32 0·52 10^{-3}(0·187416)	0·13 0·24 0·39 10^{-3}(0·332788)	0·11 0·19 0·31 10^{-3}(0·519664)	0·07 0·13 0·20 10^{-2}(0·116871)	0·06 0·10 0·15 10^{-2}(0·207913)	2,500
3,000	0·78 0·97 — 10^{-6}(0·3720)	0·70 0·93 0·98 10^{-5}(0·21007)	0·42 0·75 0·96 10^{-4}(0·211656)	0·22 0·40 0·65 10^{-4}(0·832437)	0·15 0·26 0·43 10^{-3}(0·186968)	0·11 0·20 0·32 10^{-3}(0·332066)	0·09 0·16 0·26 10^{-3}(0·518686)	0·06 0·11 0·17 10^{-2}(0·116737)	0·05 0·08 0·13 10^{-2}(0·207336)	3,000
5,000	0·78 0·97 — 10^{-7}(0·2365)	0·67 0·93 0·98 10^{-5}(0·10856)	0·26 0·48 0·77 10^{-4}(0·207426)	0·13 0·24 0·39 10^{-4}(0·827759)	0·09 0·16 0·26 10^{-3}(0·186156)	0·07 0·12 0·19 10^{-3}(0·330874)	0·06 0·10 0·15 10^{-3}(0·517561)	0·04 0·06 0·10 10^{-2}(0·116535)	0·03 0·05 0·07 10^{-2}(0·206978)	5,000
10,000	0·77 0·97 — 10^{-6}(0·1250)	0·50 0·87 0·99 10^{-5}(0·13126)	0·13 0·24 0·39 10^{-4}(0·206413)	0·07 0·12 0·19 10^{-4}(0·825366)	0·05 0·08 0·13 10^{-3}(0·185690)	0·04 0·06 0·09 10^{-3}(0·330100)	0·03 0·05 0·07 10^{-3}(0·516391)	0·02 0·03 0·05 10^{-2}(0·116288)	0·02 0·02 0·04 10^{-2}(0·206933)	10,000
50,000	0·51 0·92 — 10^{-7}(0·518)	0·11 0·19 0·31 10^{-6}(0·12874)	0·03 0·05 0·07 10^{-4}(0·206220)	0·02 0·02 0·04 10^{-4}(0·838427)						50,000
100,000	0·26 0·48 0·78 10^{-7}(0·515)	0·06 0·10 0·15 10^{-6}(0·12886)	0·02 0·02 0·04 10^{-4}(0·205566)	0·01 0·01 0·02 10^{-4}(0·838262)						100,000

* Key: β, lower 80% efficiency limit,   β°, optimal value of β,
β̄, upper 80% efficiency limit,   σ²(α̂), minimum asymptotic variance of α̂.

Should a more complicated design be used, the calculation of the maximum likelihood estimates would be much more difficult and involved. It must be pointed out that if there are serious doubts about the applicability of the exponential model, the biologist cannot rely on a design giving each test animal the same nominal dose.)

The computations and figures in this paper were prepared under the supervision of Aloise Askin and Gladys Garabedian.

MATHEMATICAL APPENDIX

Allocation of doses

Denote by β the fraction of the bacterial material reserved for the dosage part of the experiment. Let f_i be the proportion administered to the ith animal. Then

$$\beta = \sum_{i=1}^{s} f_i$$

and the nominal dose received by ith animal is $f_i \lambda$.

Since our design is to be locally optimal, we shall define the optimal design as that one which provides the maximum information relative to α (the information matrix defined by Fisher is given by

$$X = -\mathcal{E}\left(\frac{\partial^2 \log f(x, \theta)}{\partial \theta_i \partial \theta_j}\right) = \mathcal{E}\left(\frac{\partial \log f(x, \theta)}{\partial \theta_i}\frac{\partial \log f(x, \theta)}{\partial \theta_j}\right),$$

where f is the probability density and $\theta = (\alpha, \lambda)$ is the vector parameter to be estimated). Since the random variable Z (the number of bacterial colonies developing in a culture medium) observed during the plate count has a Poisson law of probability with expectation $(1-\beta)\lambda$, it follows that

$$X = \begin{pmatrix} 0 & 0 \\ 0 & \dfrac{1-\beta}{\lambda} \end{pmatrix} + \begin{pmatrix} \lambda^2 & \alpha\lambda \\ \alpha\lambda & x^2 \end{pmatrix} \sum_{i=1}^{s} \frac{f_i^2 p_i}{1-p_i}, \tag{6}$$

where $p_i = e^{-f_i \alpha \lambda}$.

For the moment let us assume the following properties of the function

$$g(u) = \frac{u^2 e^{-u}}{1-e^{-u}} \tag{7}$$

for positive u.

LEMMA 1. (a) The equation $g'(u) = 0$ has a unique positive root at $u \approx 1\cdot 6$ (the corresponding value of e^{-u} is $0\cdot 8$), and (b), $g(u)$ is concave for $0 \leqslant u \leqslant 3\cdot 08$.

It follows immediately from (a) and (1) that for an optimal design all $f_i \alpha \lambda$ should be less than $1\cdot 6$. From (b) it follows that if $0 \leqslant u_i \leqslant 3$ for all i

$$g\left(\frac{1}{s}\sum_{i=1}^{s} u_i\right) \geqslant \frac{1}{s}\sum_{i=1}^{s} g(u_i), \tag{8}$$

and therefore we have:

THEOREM 1. For the model considered here, all the animals should be given the same dose.
It remains only to prove the lemma and to determine the optimal β.

Proof of Lemma. Part (a): .

$$g'(u) = \frac{2u\,e^{-u}}{1-e^{-u}} - \frac{u^2 e^{-u}}{(1-e^{-u})^2}. \tag{9}$$

It suffices to show that the equation

$$(1-e^{-u}) - \tfrac{1}{2}u = 0, \tag{10}$$

has a unique positive root. But this follows from the fact that the expression on the left is concave and vanishes at $u = 0$.

Part (b):

$$g''(u) = (e^u - 1)^{-3} [e^{2u} (u^2 - 4u + 2) + e^u (u^2 + 4u - 4) + 2].$$

Expanding e^{2u} and e^u in a Taylor series we get

$$g''(u) = (e^u - 1)^{-3} \sum_{j=2}^{\infty} \frac{a_j u^j}{j!},$$

with $a_j = j(j-9) 2^{j-2} + 4(2^{j-2} - 1) + j^2$, so that $a_j < 0$ if $j = 2, 3, ..., 8$ and $a_j > 0$ if $j = 9, 10, ...$. From this it follows that $g''(u)$ has only one positive root (which is computed to be approximately 3·086) and (b) follows.

Suppose now that each animal is given the nominal dose $\beta\lambda/s$. The maximum-likelihood estimates of λ and α are

$$\hat{\lambda} = \frac{Z}{1-\beta} \tag{11}$$

and

$$\hat{\alpha} = \frac{-s}{\beta\hat{\lambda}} \log \hat{p} = \frac{-s}{Z} \frac{1-\beta}{\beta} \log \hat{p}, \tag{12}$$

where Z is the number of colonies developed during the plate count and $\hat{p}$ is the proportion of the animals which did not respond.

Studying $\hat{\alpha}$ directly or applying (6) it follows that the asymptotic variance of $\hat{\alpha}$ is

$$\sigma^2(\hat{\alpha}) = \frac{\alpha^2}{\lambda(1-\beta)} + \frac{s}{\beta^2\lambda^2} \frac{(1-p)}{p}, \quad p = e^{-\beta\lambda\alpha/s}. \tag{13}$$

The optimal value of β can be tabulated as a function of α and λ. Hence designs of this type can only be hoped to have local optimal properties. However, in many situations some *a priori* knowledge of λ, α is available which should enable one to get a fairly efficient design.

Examining $\sigma^2(\hat{\alpha})$ as a function of β we see that for small values of $\alpha\lambda/s$

$$\sigma^2(\hat{\alpha}) \approx \frac{s}{\beta^2\lambda^2} \left[\frac{\beta\lambda}{s} \alpha \right] + \frac{\alpha^2}{\lambda(1-\beta)} = \frac{\alpha}{\lambda} \left\{ \frac{1}{\beta} + \frac{\alpha}{1-\beta} \right\},$$

which is a minimum for $\beta = 1/(1+\sqrt{\alpha})$.

For large values of λ there should be enough material to give each animal a dose of approximately ED_{80}, which is the optimal single dose as can be determined again from the information function. If this is the case, setting

$$\beta\lambda\alpha/s = 1·6$$

will yield an approximately optimal value for β when λ is large. For an overall approximation to optimal β we propose setting

$$\beta = \beta^* = \min \left[\frac{1}{1+\sqrt{\alpha}}, \frac{(1·6)s}{\alpha\lambda} \right].$$

Table 1 gives, for various values of α, λ, the actual value of β minimizing $\sigma^2(\hat{z})$ for $s = 30$, the upper and lower limits of the values of β giving at least 80 % efficiency, and the minimum value of $\sigma^2(\hat{z})$. If Table 1 is to be used for values of s other than $s = 30$ the conversion formula

$$\sigma^2(\hat{z} \mid s, \lambda, \alpha, \beta) = \gamma \sigma^2(\hat{z} \mid \gamma s, \gamma \lambda, \alpha, \beta) \tag{14}$$

is useful, where $\sigma^2(\hat{z} \mid s, \lambda, \alpha, \beta)$ is the asymptotic variance of $\hat{z}$ when s, λ, α, β are given. For example, if $s = 60$, $\lambda = 5000$, $\alpha = 0.04$ and the minimum asymptotic variance and optimal β are desired, we see from Table 1 that for $s = 30$, $\lambda = 2500$, $\alpha = 0.04$, the optimal β is $\beta = 0.47$. From equation (14) we find that

$$\sigma^2(\hat{z} \mid s = 60, \lambda = 5000, \alpha = 0.4, \beta = 0.47)$$
$$= \tfrac{1}{2}\sigma^2(\hat{z} \mid s = 30, \lambda = 2500, \alpha = 0.4, \beta = 0.47)$$
$$= 10^{-4}(0.835757).$$

The optimal value of β is 0.47 and the upper and lower 80 % efficiency limits on β are 0.77 and 0.26, respectively.

If $\beta = \beta^*$ is used as an approximation to the optimal value of β, we see that a design of at least 80 % efficiency for all α, λ and an extremely efficient design for small λ and large λ result as can be seen in two instances in Figs. 1 and 2.

REFERENCES

CLIFTON, C. E. (1950). *Introduction to the Bacteria.* New York: McGraw Hill Book Co.

COCHRAN, W. G. (1950). Estimation of bacterial densities by means of the 'most probable number'. *Biometrics,* **6,** 105.

DRUETT, H. A. (1952). Bacterial invasion. *Nature, Lond.,* **170,** 288.

GOLDBERG, L. J. & WATKINS, H. M. S. (1952). A mathematical expression regarding mortality in the albino mouse to respiratory exposure with pathogenic micro-organisms. *Bact. Proc. Amer. Soc. Bact.* p. 74.

HALEY, K. D. C. (1953). Estimation of the dosage mortality relationship when the dose is subject to error. Unpublished Ph.D. dissertation, *Stanford University Library.*

PETO, S. (1953). A dose-response equation for the invasion of micro-organisms. *Biometrics,* **9,** 320.

WORCESTER, J. (1954). How many organisms? *Biometrics,* **10,** 227.

AN EXACT TEST FOR CORRELATION BETWEEN TIME SERIES

By E. J. HANNAN

Australian National University, Canberra, A.C.T.

1. Introduction

In a previous paper, Hannan (1955) in the process of obtaining an exact test for the serial correlation in the residuals from a regression, an estimator of the regression coefficient was also derived whose distributional properties, when the residual is a simple Gaussian Markoff process, are the same as those of the estimates of the regression coefficient in the classic cases. It was then shown that when the regressor is also generated by a simple Markoff process the regression coefficient will, for certain values of the serial correlation of the residual and the regressor, be a more efficient estimator than that obtained by straight-forward least squares. It follows that in these cases the test of significance of the regression coefficient provides a test of the correlation between the two series which is asymptotically more powerful than the approximate test based on the sample correlation coefficient.

Alternative approximate tests for correlation between two series, in the form of partial correlations with the effects of lagged values of one or both variates removed, were proposed by Quenouille (1949). In the next section of this paper the efficiencies of Quenouille's tests are compared with those of the two above-mentioned tests when the residuals from the regression come from a simple Gaussian Markoff process.

In later sections the efficiencies of these tests are considered in other cases, some of which, from the point of view of a test for correlation, appear more interesting.

2. The exact test when the residuals from the regression come from a simple Markoff process independent of the regressor process

We will consider the regression

$$y_t = \alpha + \beta x_t + \epsilon_t \quad (t = 1, ..., n), \tag{1}$$

where (1) $\epsilon_t = \rho_1 \epsilon_{t-1} + \eta_t$ and η_t is $N\{0, \sigma_1(1 - \rho_1^2)^{\frac{1}{2}}\}$, (2) ϵ_t is independent of x_s for all t and s. In a previous paper (Hannan, 1955) it was shown that an estimator of β could be obtained by considering the regression

$$y_{2t} = \alpha\left(1 - \frac{2\rho_1}{1 + \rho_1^2}\right) + \frac{\rho_1}{1 + \rho_1^2}(y_{2t-1} + y_{2t+1}) + \beta x_{2t}$$

$$- \frac{\rho_1}{1 + \rho_1^2}\beta(x_{2t-1} + x_{2t+1}) + \zeta_t \quad (t = 1, ..., [\tfrac{1}{2}(n-1)]).$$

Here ζ_t comes from a process of independent random variates with zero mean and variance $\sigma_1^2(1 - \rho_1^2)/(1 + \rho_1^2)$ and $[\tfrac{1}{2}(n-1)]$ is the greatest integer less than or equal to $\tfrac{1}{2}(n-1)$. The estimator, b_1, of β obtained from the sample regression coefficient of x_{2t} will have the usual properties of least-squares estimates on the normal case so that the corresponding

partial correlation between y_2 and x_2 will provide an exact test of the hypothesis that the two series are uncorrelated.

When x_t also comes from a simple Markoff process, with parameter ρ_2, ($\rho_2 < 1$), the variance of the estimator, b_1, was shown to be asymptotically

$$\frac{2\sigma_1^2(1-\rho_1^2)\,(1+\rho_2^2)}{n\sigma_2^2(1+\rho_1^2)\,(1-\rho_2^2)}.$$

Here σ_2^2 is the variance of x_t.

It is well known (Wold, 1953, p. 211) that the variance of the straightforward least-squares estimate, b_2, of β in the regression (1) is, asymptotically,

$$\frac{\sigma_1^2(1+\rho_1\rho_2)}{n\sigma_2^2(1-\rho_1\rho_2)}.$$

Alternative tests for correlation between two serially correlated series were proposed by Quenouille (1949). In the case of correlation between two Markoff processes he recommended the use of

$$r(x_2y_2\,|\,x_1),\quad r(x_2y_2\,|\,y_1)\quad\text{or}\quad r(x_2y_2\,|\,x_1y_1),$$

where $r(x_2y_2\,|\,x_1)$, for example, is the partial correlation between x_t and y_t with the effects of x_{t-1} removed. Asymptotically such statistics will have variances n^{-1}, when there is no correlation between the two series, since at least one of the series of residuals which are being correlated will approach independence. Quenouille showed, by a sampling experiment, that when the two series are uncorrelated the variances are approximately n^{-1}, even for small samples, and the bias is small. The best statistic, on these grounds, appeared to be $r(x_2y_2\,|\,x_1y_1)$. However, the efficiency of the statistics, as test statistics, depends not only on their variance on the null hypothesis but also on their distribution when the null hypothesis is not true. In particular, their expected values are relevant. A criterion of the asymptotic efficiency of a test based on a statistic t, of an hypothesis specified by a parameter value θ_0 is provided by the limit as $n \to \infty$ of the quantity (Stuart, 1954; Mood, 1954)

$$\left\{\frac{[\partial/\partial\theta\,\mathscr{E}(t)]^2}{\text{var}\,(t)}\right\}_{\theta=\theta_0}. \tag{2}$$

The limit of the ratio of these quantities for two statistics t_1 and t_2 has been called (Pitman, 1948) the asymptotic relative efficiency of the tests and may be denoted by $E(t_1,t_2\,|\,\theta_0)$ (where the expression (2) for t_1 is in the numerator).

As mentioned, Quenouille was considering the correlation between two Markoff processes when he proposed his statistics. In the present case y_t will not be Markovian if $\rho_1 \neq \rho_2$ and $\beta \neq 0$. It is, however, of interest to examine Quenouille's statistic in the present case also. (It is unlikely that, in practice, partial correlations of higher order than $r(x_2y_2\,|\,x_1y_1)$ will be used.)

We have
$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\beta}\mathscr{E}(r(x_2y_2\,|\,x_1))\right\}_{\beta=0} = \lim_{n\to\infty}\text{cov}\,\{r(x_2y_2\,|\,x_1),\,\mathbf{y}'\mathbf{\Gamma}_1^{-1}\mathbf{x}\},$$

when β is zero (so that the y_t and x_t processes are independent). Here $\mathbf{y}$ and $\mathbf{x}$ are vectors of the n elements y_t and x_t and $\mathbf{\Gamma}_1$ is the covariance matrix of the residuals ϵ_t. The differentiation under the integral sign is justified by the uniform convergence of the resulting integral.

The limit of the covariance can be evaluated by straightforward methods (making use of an obvious extension of the theorem quoted in Cramer, 1946, p. 353) and gives

$$\lim_{n \to \infty} \left\{ \frac{\partial}{\partial \beta} \mathscr{E}(r(x_2 y_2 \mid x_1)) \right\}_{\beta = 0} = \frac{\sigma_2 (1 - \rho_2^2)^{\frac{1}{2}}}{\sigma_1}.$$

Similarly

$$\lim_{n \to \infty} \left\{ \frac{\partial}{\partial \beta} \mathscr{E}(r(x_2 y_2 \mid x_1 y_1)) \right\}_{\beta = 0} = \frac{\sigma_2}{\sigma_1} \left(\frac{1 - \rho_2^2}{1 - \rho_1^2} \right)^{\frac{1}{2}}.$$

Since these two statistics have the same limiting variance on the null hypothesis $r(x_2 y_2 \mid x_1 y_1)$ is always asymptotically the more efficient.

We now have

$$E(b_1, b_2 \mid \beta = 0) = \frac{(1 + \rho_1 \rho_2)(1 + \rho_1^2)(1 - \rho_2^2)}{2(1 - \rho_1 \rho_2)(1 - \rho_1^2)(1 + \rho_2^2)},$$

$$E\{r(x_2 y_2 \mid x_1 y_1), b_2 \mid \beta = 0\} = \frac{(1 + \rho_1 \rho_2)(1 - \rho_2^2)}{(1 - \rho_1 \rho_2)(1 - \rho_1^2)},$$

$$E\{r(x_2 y_2 \mid x_1 y_1), b_1 \mid \beta = 0\} = 2 \left(\frac{1 + \rho_2^2}{1 + \rho_1^2} \right).$$

The first of these ratios was shown for certain values of ρ_1 and ρ_2 in Table 3 in Hannan (1955).[*] The second ratio is shown for the same values of ρ_1 and ρ_2 in Table 1.

Table 1. $E(r(x_2 y_2 \mid x_1 y_1), b_2 \mid \beta = 0)$

ρ_2 \ ρ_1	$-0\cdot8$	$-0\cdot6$	$-0\cdot4$	$-0\cdot2$	0	$0\cdot2$	$0\cdot4$	$0\cdot6$	$0\cdot8$
0	2·78	1·56	1·19	1·04	1	1·04	1·19	1·56	2·78
0·2	1·93	1·18	0·97	0·92	0·96	1·08	1·34	1·91	3·68
0·4	1·20	0·80	0·72	0·75	0·84	1·03	1·38	2·14	4·53
0·6	0·62	0·47	0·47	0·52	0·64	0·85	1·24	2·13	5·06
0·8	0·22	0·20	0·22	0·27	0·36	0·52	0·83	1·60	4·56

It is evident that, while b_1 will be asymptotically more powerful than b_2 when ρ_1 is sufficiently high it is always asymptotically less efficient than $r(x_2 y_2 \mid x_1 y_1)$. It has the advantage, however, of giving an exact test.

It is easily shown that when x_t is Markovian $- \mathscr{E} \left(\frac{\partial^2 \log L}{\partial \beta^2} \right)_{\beta = 0}$ tends to

$$\{ n \sigma_2^2 (1 + \rho_1^2 - 2 \rho_1 \rho_2) \} / \{ \sigma_1^2 (1 - \rho_1^2) \}$$

as n increases; while the quantities $- \mathscr{E} \left(\frac{\partial^2 \log L}{\partial \beta \partial \sigma_1^2} \right)_{\beta = 0}$ and $- \mathscr{E} \left(\frac{\partial^2 \log L}{\partial \beta \partial \rho_1} \right)_{\beta = 0}$ tend to zero.

Here L is the likelihood function. It follows (Whittle, 1953) that the variance of the maximum-likelihood estimator of β is, asymptotically, when $\beta = 0$

$$\frac{\sigma_1^2 (1 - \rho_1^2)}{n \sigma_2^2 (1 + \rho_1^2 - 2 \rho_1 \rho_2)}.$$

None of the statistics considered in this section are, therefore, asymptotically fully efficient in general, though $r(x_2 y_2 \mid x_1 y_1)$ is asymptotically fully efficient when $\rho_1 = \rho_2$.

[*] There ρ and ρ_1 were used in place of ρ_1 and ρ_2 respectively.

One can, of course, suggest tests which are asymptotically fully efficient and which do not require the solution of non-linear systems of equations as the maximum-likelihood estimator does. Such a test will be obtained by using a consistent estimator of the serial correlation of the residuals to transform the regression equation to a form with an approximately random residual (Cochrane & Orcutt, 1949). Apart from the fact that the criterion of asymptotic relative efficiency is only suggestive of the relative power of tests in small samples it has an added deficiency when applied to tests which are not exact, for it does not take into account the effect of the deviation of the significance point used from the true value appropriate to the level of significance required. In the case of Quenouille's tests his sampling experiments seem to show that on the null hypothesis the mean and variance of his statistics in small samples are near to their theoretical values so that the significance point used will be approximately correct.

Example 1. Two series of seventy observations were chosen from series 7 in Kendall (1949), the last term in one series being sufficiently far from the first in the next to make the two series effectively independent. If the two series are represented by x_t and z_t, a third series which was formed can be represented by

$$y_t = 0 \cdot 8 x_t + z_t.$$

The commencing term in the series x_t is no. 191 and in z_t no. 316.

Table 2

Statistic	Sample value	Population value	Test statistic
b_1	0·994	0·8	$t_{30} = 5 \cdot 27$**
b_2	0·417	0·8	$r_{30} = 0 \cdot 316$
$r(x_2 y_2 \mid x_1)$	0·316	0·329	$r_{67} = 0 \cdot 361$**
$r(x_2 y_2 \mid x_1 y_1)$	0·684	0·625	$r_{66} = 0 \cdot 684$**

** Highly significant.

The three series now are Gaussian Markoff processes with first serial correlation equalling 0·9.

The details of the various tests are given in Table 2.

Here t_{30} indicates Student's t with 30 degrees of freedom, while r_m indicates a correlation coefficient with m degrees of freedom.

The observed first serial correlation coefficients were

$$r_{1,x} = 0 \cdot 755, \quad r_{1,y} = 0 \cdot 739.$$

The degrees of freedom appropriate to r_{30} were calculated from the formula

$$n' = 70 \frac{1 - (0 \cdot 755)(0 \cdot 739)}{1 + (0 \cdot 755)(0 \cdot 739)} \approx 20.$$

The statistic r_{20} is not significant at the 5 % point.

3. An exact test for correlation between two simple Markoff processes

Let

$$(y_t - \mu_1) = \rho_1(y_{t-1} - \mu_1) + \epsilon_t,$$

$$(x_t - \mu_2) = \rho_2(x_{t-1} - \mu_2) + \eta_t,$$

be two simple Markoff processes with normally distributed disturbances having zero means and variances $\sigma_1^2(1 - \rho_1^2)$ and $\sigma_2^2(1 - \rho_2^2)$.

The correlation between these processes can be prescribed in terms of the correlation properties of the residuals. Since these two residuals come from processes of independent random variates this will be achieved by saying for what lag, if any, the two residuals are correlated and what these correlations are.

We will consider the case where the only non-zero correlation is between ϵ_t and η_{t-m} and the series have been so lagged relative to each other that m is zero.

If the correlation between ϵ_t and η_t is ρ, then the correlation between x_t and y_s is

$$t \geqslant s: \quad \frac{\rho\{(1-\rho_1^2)(1-\rho_2^2)\}^{\frac{1}{2}}}{(1-\rho_1\rho_2)} \rho_2^{t-s} = \alpha \rho_2^{t-s},$$

$$t \leqslant s: \qquad\qquad = \alpha \rho_1^{s-t}.$$

The two series y_t and x_t will be jointly normally distributed for every t so that

$$(y_t - \mu_1) = \alpha \frac{\sigma_1}{\sigma_2}(x_t - \mu_2) + \zeta_t,$$

where ζ_t is independent of x_t and has mean zero and variance $\sigma_1^2(1 - \alpha^2)$.

Then a simple calculation shows that $\mathcal{E}(\zeta_t \zeta_{t-s}) = \sigma_1^2(1 - \alpha^2)\rho_1^s$. It follows, since ζ_t is normally distributed, that it is generated by a simple Gaussian Markoff process with parameter ρ_1 (Doob, 1953, p. 233). The disturbance θ_t in the process $\zeta_t = \rho_1\zeta_{t-1} + \theta_t$ is

$$\theta_t = \alpha \frac{\sigma_1}{\sigma_2}(\rho_1 - \rho_2)(x_{t-1} - \mu_2) + \epsilon_t - \alpha \frac{\sigma_1}{\sigma_2}\eta_t.$$

If $\rho_1 \neq \rho_2$ and $\rho \neq 0$, θ_t depends upon x_{t-1} and ζ_t is not independent of x_s for $t \neq s$. The θ_t are of course independent.

When $\rho_1 = \rho_2$ the present case is equivalent to that considered in the previous section. When $\rho_1 \neq \rho_2$ the exact test of significance of the correlation between the two series there given is still an exact test, but the conclusions as to the power of this test and the others there considered, when x_t also comes from a Markoff process, do not follow, for when $\rho \neq 0$ the condition (2) upon which the derivation depended does not hold.

The conditional distribution of the y_{2t} ($t = 1, \ldots, [\frac{1}{2}(n-1)]$), for fixed x_t ($t = 1, \ldots, n$) and y_{2t+1} ($t = 0, \ldots, [\frac{1}{2}(n-1)]$) is the product of the conditional distributions of each y_{2t} for fixed $y_{2t-1}, y_{2t+1}, x_{2t-1}, x_{2t}$ and x_{2t+1}. This follows from the fact that for fixed y_{2t-1} and x_{2t+1} the y_{2t} and x_{2t} are serially independent (Ogawara, 1951). This conditional distribution can be found

from the correlation matrix of the variates $y_{2l-1}, y_{2l-1}, x_{2l-1}, x_{2l}, x_{2l+1}, y_{2l}$ (taken in that order), which is

$$\Lambda = \begin{bmatrix} A & B \\ B' & 1 \end{bmatrix},$$

$$A = \begin{bmatrix} 1 & \rho_1^2 & \alpha & \alpha\rho_2 & \alpha\rho_2^2 \\ \rho_1^2 & 1 & \alpha\rho_1^2 & \alpha\rho_1 & \alpha \\ \alpha & \alpha\rho_1^2 & 1 & \rho_2 & \rho_2^2 \\ \alpha\rho_2 & \alpha\rho_1 & \rho_2 & 1 & \rho_2 \\ \alpha\rho_2^2 & \alpha & \rho_2^2 & \rho_2 & 1 \end{bmatrix},$$

$$B' = \{\rho_1 \quad \rho_1 \quad \alpha\rho_1 \quad \alpha \quad \alpha\rho_2\}.$$

Then the variate y_{2l} is conditionally distributed with mean $\sigma_1 A^{-1}BZ$ and variance $\sigma_1^2\{1 - B'A^{-1}B\}$ (Rao, 1952, p. 54).

Here

$$Z' = \left| \frac{y_{2l-1}}{\sigma_1} \quad \frac{y_{2l+1}}{\sigma_1} \quad \frac{x_{2l-1}}{\sigma_2} \quad \frac{x_{2l}}{\sigma_2} \quad \frac{x_{2l+1}}{\sigma_2} \right|.$$

The vector $(A^{-1}B)'$ is easily seen to be

$$\frac{1}{1+\rho_1^2}\left\{\rho_1 \quad \rho_1 \quad \frac{-\rho_2\alpha(1-\rho_1\rho_2)}{1-\rho_2^2} \quad \frac{\alpha(1-\rho_1^2\rho_2^2)}{1-\rho_2^2} \quad \frac{-\rho_1\alpha(1-\rho_1\rho_2)}{1-\rho_2^2}\right\}',$$

while

$$\sigma_1^2\{1 - B'A^{-1}B\} = \frac{\sigma_1^2(1-\rho^2)(1-\rho_1^2)}{1+\rho_1^2}.$$

For $\rho_1 = \rho_2$ these formulae are equivalent to those given in the previous section, when it is remembered that there the variance of the residual ϵ_t was σ_1^2, while here the variance of the corresponding residual is $\sigma_1^2(1-\rho^2)$.

For $\rho_1 \neq \rho_2$ two things are noticeable:

(1) The regression coefficient of x_{2l-1} is not equal to that for x_{2l+1} so that some of the symmetry present when $\rho_1 = \rho_2$ is now lost. This loss of symmetry is also evident from the fact that

$$\mathcal{E}\{(x_t - \mu_2)(y_{t+s} - \mu_1)\} \neq \mathcal{E}\{(x_{t+s} - \mu_2)(y_t - \mu_1)\},$$

when

$$\rho_1 \neq \rho_2 \quad \text{and} \quad s \neq 0$$

and

$$\mathcal{E}\{(y_t - \mu_1)(x_{t+s} - \mu_2) \mid y_{t+1}, y_{t-1}\} = 0 \quad s < 0,$$
$$\neq 0 \quad s > 0.$$

(2) The coefficient of x_{2l} is no longer equal to $\alpha\sigma_1/\sigma_2$ but instead to

$$\alpha\frac{\sigma_1}{\sigma_2}\left\{\frac{1-\rho_1^2\rho_2^2}{(1+\rho_1^2)(1-\rho_2^2)}\right\}.$$

The asymptotic variance of the estimator, b_1, of this last coefficient is, for $\rho = 0$,

$$\frac{2\sigma_1^2(1-\rho_1^2)(1+\rho_2^2)}{n\sigma_2^2(1+\rho_1^2)(1-\rho_2^2)},$$

so that the quantity

$$\lim_{n\to\infty}\left\{\text{var}(b_1)\middle/\left[\frac{\partial}{\partial\rho}\mathcal{E}(b_1)\right]^2\right\}_{\rho=0}$$

is

$$\frac{2(1+\rho_1^2)(1+\rho_2^2)}{n(1+\rho_1\rho_2)^2}.$$

For the ordinary correlation coefficient, r, between x_t and y_t the corresponding quantity is (Bartlett, 1935)

$$\frac{1-\rho_1^2\rho_2^2}{n(1-\rho_1^2)(1-\rho_2^2)}.$$

The ratio of the second of these quantities to the first (the asymptotic relative efficiency of the two tests) is

$$E(b_1, r \mid \rho=0) = \frac{1}{2}\frac{(1-\rho_1^2\rho_2^2)(1+\rho_1\rho_2)^2}{(1-\rho_1^4)(1-\rho_2^4)}.$$

which is evaluated for certain values of ρ_1 and ρ_2 in Table 3. For $E(b_1, r \mid \rho=0) > 1$ it may be inferred that, asymptotically at least, the test based on b_1 is superior to that based on r.

Table 3. $E(b_1, r \mid \rho=0)$

ρ_2 / ρ_1	$-0\cdot8$	$-0\cdot6$	$-0\cdot4$	$-0\cdot2$	0	$0\cdot2$	$0\cdot4$	$0\cdot6$	$0\cdot8$
0	0·85	0·57	0·51	0·50	0·50	0·50	0·51	0·57	0·85
$0\cdot2$	0·58	0·44	0·43	0·46	0·50	0·54	0·60	0·71	1·11
$0\cdot4$	0·36	0·32	0·36	0·43	0·51	0·60	0·69	0·85	1·36
$0\cdot6$	0·20	0·24	0·32	0·44	0·57	0·71	0·85	1·06	1·64
$0\cdot8$	0·11	0·20	0·36	0·58	0·85	1·11	1·36	1·64	2·28

As in § 2 we have

$$\lim_{n \to \infty}\left\{\frac{\partial}{\partial\rho}\mathscr{E}(r(x_2y_2 \mid x_1))\right\}_{\rho=0} = \lim_{n\to\infty}\text{cov}\,\{r(x_2y_2 \mid x_1), \mathbf{y}'\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_3\boldsymbol{\Gamma}_2^{-1}\mathbf{x}\},$$

the differentiation having been carried out under the integral sign and $\mathbf{y}$ now being independent of $\mathbf{x}$. Here $\boldsymbol{\Gamma}_1$ is the covariance matrix of the y_t, $\boldsymbol{\Gamma}_2$ is the covariance matrix of the x_t and

$$\boldsymbol{\Gamma}_3 = \frac{1}{\rho}[\mathscr{E}\{(y_i-\mu_1)(x_j-\mu_2)\}].$$

The limit of the covariance can be evaluated in the same manner as in § 2 and gives

$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\rho}\mathscr{E}(r(x_2y_2 \mid x_1))\right\} = (1-\rho_1^2)^{\frac{1}{2}},$$

$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\rho}\mathscr{E}(r(x_2y_2 \mid y_1))\right\} = (1-\rho_2^2)^{\frac{1}{2}},$$

$$\lim_{n\to\infty}\left\{\frac{\partial}{\partial\rho}\mathscr{E}(r(x_2y_2 \mid x_1y_1))\right\} = 1.$$

Clearly once more $r(x_2y_2 \mid x_1y_1)$ provides the most efficient test statistic.

It can be shown that, for $\rho = 0$, the quantities $-\mathscr{E}\left(\frac{\partial^2\log L}{\partial\rho\partial\theta_j}\right)_{\rho=0}$ are asymptotically zero. Here L is the likelihood function and the θ_j ($j = 1, 2, 3, 4$) are the remaining parameters in that function. At $\rho = 0$, $-\mathscr{E}\left(\frac{\partial^2\log L}{\partial\rho^2}\right) = n$. It therefore appears (Whittle, 1953) that the asymptotic variance of the maximum-likelihood estimator of ρ is n^{-1} when ρ is zero. Since

this estimator will be consistent both r and b_1 are asymptotically inefficient as test statistics when ρ_1 and ρ_2 are not zero, while $r(x_2y_2, x_1y_1)$ is asymptotically fully efficient.

When ρ_1 and ρ_2 are high and of the same sign b_1 will give an exact test of the null hypothesis which is more efficient than the test based on r. However, the test based on b_1 is always much less efficient than that based on $r(x_2y_2, x_1y_1)$.

The computational burden of the exact test can of course be reduced by omitting one or both of $x_{2,1}$ and $x_{2,1}$ from the regression and the test will still be exact. It can be shown, however, that the test will be asymptotically less efficient. Another alternative is to use $(x_{2,1} + x_{2,1})$ in place of $x_{2,1}$ and $x_{2,1}$. It can then be shown that the asymptotic efficiency of the test based on the coefficient of x_{2} is unchanged. In finite samples, however, the power of the test may be adversely affected.

The following example illustrates the use of some of the tests mentioned in this section.

Example 2. In Moran (1952) the correlations between the records of the numbers of certain game birds shot over a period of years were considered. The series for caper and ptarmigan are considered in this example.

The application of the test given in Quenouille (1947a) shows that both series are well represented by a simple Markoff process.

The results of the various tests discussed above are shown in Table 4. Here r_m indicates a correlation coefficient with m degrees of freedom.

Table 4

Statistic	Sample value	Test statistic
r	0·417	$r_{34} = 0\cdot417^*$
b_1	0·444	$t_{31} = 2\cdot484^*$
$r(x_2y_2\,\vert\,x_1y_1)$	0·387	$r_{60} = 0\cdot387^{***}$

* Significant at 5 % point.   *** Significant at 0·1 % point.

The two observed serial correlations were 0·592 and 0·417 for caper and ptarmigan respectively, and these were adjusted up to 0·65 and 0·55 by Moran to allow for bias. Using these last values it appears from Table 3 that r and b_1 should give tests of roughly the same efficiency while $r(x_2y_2 \,\vert\, x_1y_1)$ should be approximately twice as efficient as either.

4. The case where the regressor process is a second-order autoregressive process

The exact tests for correlation between two series which have been presented in this paper remain exact provided that the residual from the regression of one variate on the other is a Markoff process. The efficiency of the tests has been examined in the following cases:

(1) The regressor process is a Markoff process wholly independent of the residual. In particular this covers the case where the regressand is also a Markoff process with the same parameter as the regressor and the residual.

(2) The regressand and regressor processes are both Markovian. When the two parameters are equal this case is equivalent to the particular case mentioned under (1) above.

If neither of the two processes being correlated can reasonably be said to be Markovian

then strictly, the exact tests presented here are not applicable. For the result are really Markovian only if there is in truth a relation between the series. (One might apply them as an approximation to an exact test of course.)

A case which may arise in practice is that in which one process appears Markovian while the other is an autoregressive process of higher order. The more interesting alternative hypothesis then appears to be one in which the correlation between the series arises from a correlation between the shocks of which they are linearly composed.

Let

$$(y_t - \mu_1) = \rho_1(y_{t-1} - \mu_1) + \epsilon_t$$

and

$$(x_t - \mu_2) + a(x_{t-1} - \mu_2) + b(x_{t-2} - \mu_2) = \eta_t$$

be two processes where,

(1) $|\rho_1| < 1$ and all roots of $z^2 + az + b = 0$ are less than unity in absolute value.
(2) ϵ_t and η_t are jointly normally distributed with variances

$$\sigma_1^2(1 - \rho_1^2) \quad \text{and} \quad \sigma_2^2\{(1-b)(1+b-a^2)(1+b)^{-1}\}$$

and correlation ρ.

Table 5. $E(b_1, r \mid \rho = 0)$

b	ρ_1, ρ_2 $\rho_1 = \rho_2 = 0\cdot4$	$\rho_1 = \rho_2 = 0\cdot6$	$\rho_1 = \rho_2 = 0\cdot8$	$\rho_1 = 0\cdot6$ $\rho_2 = 0\cdot8$	$\rho_1 = 0\cdot8$ $\rho_2 = 0\cdot6$
$-0\cdot6$	$0\cdot82$	$1\cdot42$	$3\cdot45$	$2\cdot73$	$1\cdot82$
$-0\cdot4$	$0\cdot77$	$1\cdot30$	$3\cdot04$	$2\cdot33$	$1\cdot79$
$-0\cdot2$	$0\cdot73$	$1\cdot18$	$2\cdot65$	$1\cdot97$	$1\cdot73$
0	$0\cdot69$	$1\cdot06$	$2\cdot28$	$1\cdot64$	$1\cdot64$
$0\cdot2$	$0\cdot65$	$0\cdot95$	$1\cdot94$	$1\cdot36$	$1\cdot56$
$0\cdot4$	$0\cdot62$	$0\cdot86$	$1\cdot64$	$1\cdot12$	$1\cdot42$
$0\cdot6$	$0\cdot59$	$0\cdot79$	$1\cdot40$	$0\cdot95$	$1\cdot33$

Then by straightforward methods it can be shown that

$$E(b_1, r \mid \rho = 0) = \frac{(1 + \rho_1\rho_2 - b)^2\{1 + \rho_1\rho_2 - b(\rho_1^2 + \rho_1\rho_2)\}\{1 - \rho_1\rho_2 - b(\rho_1\rho_2 - \rho_1^2)\}}{2(1 - \rho_1^4)(1 - \rho_2^2)(1 - b)\{1 + \rho_2^2 - b(1 - \rho_2^2)\}}.$$

Here ρ_2 is the first serial correlation of the x_t process.

At $b = 0$ this is equal to the quantity tabulated above. The ratio decreases with b, for most values of ρ_1 and ρ_2, at least in the neighbourhood of $b = 0$. For sufficiently large values of ρ_1 and ρ_2 it will be greater than unity. The ratio is tabulated in Table 5 for certain values of ρ_1, ρ_2 and b.

It is clear that the statistic $r(x_3y_3, x_1x_2y_2)$ will now give a test which is asymptotically fully efficient. The asymptotic efficiency of the statistic $r(x_2y_2 \mid x_1y_1)$ can be fairly easily evaluated and proves to be

$$E\{r(x_2y_2 \mid x_1y_1), r(x_3y_3 \mid x_1x_2y_2) \mid \rho = 0\} = (1 - b^2).$$

The use of $r(x_2y_2 \mid x_1y_1)$ will then result in little loss of information unless b is fairly large.

Finally

$$E\{b_1, r(x_2y_2 \mid x_1y_1)\} = \frac{(1 - b + \rho_1\rho_2)^2}{2(1 + \rho_1^2)(1 - b)\{1 - b + \rho_2^2(1 + b)\}}.$$

Thus, for that part of the range of values of the parameters investigated, though it becomes infinite as b tends to unity.

Example 3. The data for grouse and chicks grouse from Moran (1952) may be used for illustration. The blackgrouse mating separates to G. Markovian, but we find grouse mating separate to be the grouse series ρ 0·54. The data were combined to [for grouse ρ b_1 and the Markovian 7·43. Using $b_1 = 0.8416, r = 0.77$ and $r_1 = 0.77$ (Fisher),

$$E(b_1, r \mid \rho = 0) = 0.02, \quad E(b_1, r(x_2 y_2 \mid x_1 y_1 \mid \rho = 0) = 0.47$$

The tests are shown in Table 6.

Table 6

Statistic	Sample value	Test statistic
b_1	0·0754	$t_5 = 2.40^*$
r	0·614	$t_9 = 3.84^*$
$r(x_2 y_2 \mid x_1 y_1)$	0·377	$t_5 = 7.12^{**}$

* Significant at 5 per cent.   ** Significant at 1 per cent.

5. GENERALIZATION

The preceding exact tests can of course be generalized to multiple correlation and analogous results can be expected.

A further generalization of both the simple and multiple correlation tests can be obtained when one process is an autoregressive process of higher order than the first. For a process of order z, however, only $m(h+1)^{-1}$ observations will be used in the correlation, where n is the number of observations available, the remainder being used to reduce the series to independence in time. The serial correlations would have to be very high indeed before the test began to compare favourably with one based on the ordinary correlation (or multiple correlation) coefficient. The computational burden involved in the production of the partial correlation (or partial multiple correlation) coefficient will also be great.

6. SUMMARY

An exact test of the correlation between two series has been obtained which can be applied whenever one of the two series is Markovian. If the two series are x_t and y_t and y_t is Markovian the test statistic, b_1, is the partial correlation between x_2 and y_2 when the effects of $(y_{2-1} + y_{2-1}) x_{2-1}$ and x_{2-1} have been removed. The asymptotic efficiency of this statistic is compared with that of the ordinary correlation coefficient r, between the two series and the statistic $r(x_2 y_2 \mid x_1 y_1)$ suggested by Quenouille (1949) under the following conditions:

(a) The residual process from the regression of y_t on x_t is independent of the x_t process and comes from a Gaussian Markoff process.

(b) The two series x_t and y_t are Markovian and are correlated through a correlation between the (Gaussian) errors in the two processes.

(c) As in (b) but with x_t generated by a second order autoregressive process.

While in all three cases the statistic b_1 leads to an asymptotically more efficient test than r for high serial correlation of the residual process in the regression of y_t on x_t, the statistic

$r(x_2 y_2 \mid x_1 y_1)$ is always asymptotically more efficient than b_1 with the exception of some cases, under (c), where the first partial correlation of the x_t process is high and positive.

I should like to thank a referee of this paper for pointing out the importance of Quenouille's statistic.

REFERENCES

BARTLETT, M. S. (1935). *J. R. Statist. Soc.* **98**, 536.
COCHRANE, D. & ORCUTT, G. H. (1949). *J. Amer. Statist. Ass.* **44**, 32.
CRAMER, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.
DOOB, J. L. (1953). *Stochastic Processes*. New York: John Wiley and Sons.
HANNAN, E. J. (1955). *Biometrika*, **42**, 133.
KENDALL, M. G. (1949). *Biometrika*, **36**, 267.
MOOD, A. M. (1954). *Ann. Math. Statist.* **25**, 514.
MORAN, P. A. P. (1952). *J. Animal Ecol.* **21**, 154.
OGAWARA, M. (1951). *Ann. Math. Statist.* **22**, 115.
PITMAN, E. J. G. (1948). Lecture notes on non-parametric inference. University of North Carolina.
QUENOUILLE, M. H. (1947a). *J. R. Statist. Soc.* **60**, 123.
QUENOUILLE, M. H. (1947b). *Biometrika*, **34**, 365.
QUENOUILLE, M. H. (1949). *J. R. Statist. Soc.* B, **11**, 68.
RAO, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: John Wiley and Sons.
STUART, A. (1954). *J. Amer. Statist. Ass.* **49**, 147.
WHITTLE, P. (1953). *J. R. Statist. Soc.* B, **15**, 125.
WOLD, H. (1953). *Demand Analysis*. New York: John Wiley and Sons.

SERIAL CORRELATION IN REGRESSION ANALYSIS. I†

By G. S. WATSON

Australian National University, Canberra, A.C.T.

1. Introduction

The procedure for linear regression analysis when the errors have a multivariate normal joint distribution with covariance matrix $\sigma^2\alpha$, where α is a *given* matrix are well known (see, for example, Aitken, 1934–5). However, in many practical applications, the matrix α is not completely known. In the analysis of ordered observations, the errors are often serially correlated. Even if the errors are assumed to be generated by a stationary stochastic process of the moving-average or autoregressive type of lower order, the actual selection of the process and its parameters is a difficult matter, requiring samples of a size not often available. It is therefore certain that regression analyses will often be made using the wrong matrix α.

Recently Hannan (1955) has given a method which, while not fully efficient, leads to exact tests when the error process is Markov with unknown ρ. This is an advance on previous writers (e.g. Wold, 1949; Cochrane & Orcutt, 1949), who aimed at the estimation of the serial correlations of the errors and hence the determination of α. But the basic difficulty still remains—the error process may not be Markov. Furthermore, while the difficulties are well known, the orders of magnitudes of the effects of using the wrong matrix α have received little precise study except for the case of unequal variances in the analysis of variance (see, for example, Box, 1954).

For these reasons in Part I of this paper an examination will be made of the performance of a regression analysis based on the assumption that the error covariance matrix is $\sigma^2\gamma$ when it is, in fact, $\sigma^2\alpha$. The methods used have their origin in the papers of Durbin & Watson (1950, 1951). When $\gamma \neq \alpha$ the regression vectors play an important role in determining the efficiency of the regression coefficient estimates and the true significance levels of significance tests. The procedure adopted is to find, for fixed α and γ, the regression vectors which would make the analysis as 'bad' as possible; this leads to inequalities on the bias in the estimates of variance of the regression coefficients, on the efficiency of the estimates of the regression coefficients and on the significance points of the various t- and F-tests.

In Part II, these inequalities will be applied to cases where $\sigma^2\alpha$ and $\sigma^2\gamma$ are the covariance matrices of first- and second-order autoregressive and moving-average processes. By using approximations to these matrices, it is possible to get simple algebraic and numerical results.

We consider the linear regression of a dependent variable y on k (linearly independent) regression variables $x_1, x_2, \ldots, x_k$. The model for a sample of N observations is, in matrix form,

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{k1} \\ \vdots & & \vdots \\ x_{1N} & \cdots & x_{kN} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}, \tag{1·1}$$

† This work was carried out while the writer was a Research Officer in the Department of Applied Economics, University of Cambridge. The results of this paper and its sequel were presented at the Conference of the Econometric Society, Louvain, Belgium in September 1951. It formed part of a thesis which was issued by the Institute of Statistics, University of North Carolina, Mimeograph Series, no. 49.

or
$$y = [x_1, ..., x_k]\beta + u,$$

or
$$y = X\beta + u,$$

where the error vector u has zero expectation and

$$E(uu') = \sigma^2\alpha \quad (\alpha \text{ non-singular}). \tag{1·2}$$

σ^2, α and β are supposed unknown. If the statistician, analysing these data, decides *a priori* that the error covariance matrix is $\sigma^2\gamma$, he will essentially proceed as follows. Since γ is known it is possible to find a non-singular $N \times N$ matrix H such that

$$H\gamma H' = I. \tag{1·3}$$

Writing
$$Hy = y^*, \quad Hx_i = x_i^*, \quad Hu = u^*, \quad HX = X^*,$$

the result of applying H to equation (1·1) is the transformed regression equation

$$y^* = X^*\beta + u^*, \tag{1·4}$$

where u^* has zero expectation and

$$E(u^*u'^*) = \sigma^2 H\alpha H'. \tag{1·5}$$

The regression equation (1·4) will now be treated by the least-squares procedures.† Thus in our analysis of the effects of 'guessing' the error covariance matrix, we may without loss of generality proceed on the assumption that the least-squares procedure has been used on the model represented by equations (1·1) and (1·2) and merely replace α by $H\alpha H'$ in the final results.

On these assumptions, the estimate of β used would be

$$b = (X'X)^{-1}X'y. \tag{1·6}$$

The covariance matrix of the components of b is actually

$$\sigma^2(X'X)^{-1}X'\alpha X(X'X)^{-1}, \tag{1·7}$$

but it would be taken as

$$\sigma^2(X'X)^{-1}, \tag{1·8}$$

where σ^2 would be estimated by

$$s^2 = \frac{(y - Xb)'(y - Xb)}{N - k}. \tag{1·9}$$

The true expected value of s^2 is given by

$$E(s^2) = \frac{1}{N-k} E((u - X(X'X)^{-1}X'u)'(u - X(X'X)^{-1}X'u))$$

$$= \frac{1}{N-k} E(u'u - u'X(X'X)^{-1}X'u)$$

$$= \frac{\sigma^2}{N-k} (\operatorname{tr}\alpha - \operatorname{tr}X'\alpha X(X'X)^{-1}). \tag{1·10}$$

For arbitrary X this reduces to σ^2 only when $\alpha = I$.

† By the 'least-squares procedures' is meant not only the estimation of the regression coefficients by least squares but also the use of the variance estimates and the t- and F-tests which are appropriate to the case where the error covariance matrix really is $I\sigma^2$. We might have introduced the name 'I-procedure' to cover this. Thus when the statistician decides the error covariance matrix is $\sigma^2\gamma$ he will use the 'γ-procedure'.

2. BOUNDS ON THE BIAS OF THE VARIANCE ESTIMATES

The bias in the estimate of the covariance matrix of the regression coefficient estimates may be found from (1·7), (1·8) and (1·10) and is given by

$$\sigma^2(X'X)^{-1}\left(\frac{\operatorname{tr}\alpha - \operatorname{tr} X'\alpha X(X'X)^{-1}}{N-k}I - X'\alpha X(X'X)^{-1}\right). \tag{2·1}$$

In order to examine the magnitude of the effect on the variance of a single regression coefficient we now assume that the regression vectors form an orthonormal set, i.e.

$$X'X = I. \tag{2·2}$$

The condition (2·2) invoked to simplify (2·1) leads to a loss of generality; but, without it, the extreme value problem is very difficult. Fortunately we will not be so restricted when we come to discuss significance tests. On this assumption the estimate of β_i will be

$$b_i = x'_i y, \tag{2·3}$$

and the bias in the estimate of its variance will be

$$\sigma^2\left(\frac{\operatorname{tr}\alpha - \operatorname{tr} X'\alpha X}{N-k} - x'_i\alpha x_i\right),$$

that is,

$$\sigma^2\left(\frac{\operatorname{tr}\alpha - \sum\limits_{j=1}^{k} x'_j\alpha x_j}{N-k} - x'_i\alpha x_i\right). \tag{2·4}$$

We now require the extremes of the expression (2·4). These will be seen to follow easily from the following algebraic result: the extremes of the quadratic form $Z'\alpha Z$, where Z is a unit vector in a subspace spanned by a subset of the latent vectors of α, are the least and greatest latent roots of α associated with the subspace. Thus the lower bound of (2·4) is obtained by choosing one of the x-vectors, say x_i, to be the latent vector of α corresponding to its largest latent root and the others to be the latent vectors corresponding to the next $k-1$ largest roots. The upper bound is found by a similar choice using the smallest roots of α. More generally if h ($< k$) of the regression vectors are latent vectors of α, the remaining regression vectors lie in an $N-h$ dimensional subspace. If x_i lies in this subspace, the above results hold good if the choices are made with respect to the set of the $N-h$ latent roots associated with this subspace. If x_i is a latent vector, a trivial modification is necessary.

Thus for X subject only to (2·2)

(a) Maximum bias = $\sigma^2((\text{mean of } N-k \text{ greatest roots of } \alpha) - (\text{least root of } \alpha))$,
(b) Minimum bias = $\sigma^2((\text{mean of } N-k \text{ least roots of } \alpha) - (\text{greatest root of } \alpha))$. $\quad$ (2·5)

If the multiple σ^2 is omitted the results are in the nature of bounds on the fractional bias. It will be noticed that (2·5 (a)) is positive and (2·5 (b)) negative *whatever* α may be. It is commonly believed that presence of serial correlation makes the variance estimates deceptively small. The numerical results given in Part II show this to be the stronger tendency, but (2·5) shows that this need not always be the case. The deciding factor is seen to be the relationship of the regression vectors to the latent vectors of α. Finally, (2·5) shows how the bias must tend to zero as $\alpha \to I$.

3. Lower bound to the efficiency of the estimates

We turn now to a consideration of the efficiency of the estimates (1·6) of the regression coefficients. It is necessary to determine how this efficiency falls off as α differs more and more from I. To consider the efficiency of the estimation procedure as a whole it is convenient to introduce the generalized variance of the estimates. The generalized variance of the estimates b_i of (1·6) is defined to be the determinant of their covariance matrix (1·7). Aitken (1948) has shown that the linear estimates of β_i with the least generalized variance are those given by $(X'\alpha^{-1}X)^{-1}X'\alpha^{-1}y$; they have the covariance matrix $\sigma^2(X'\alpha^{-1}X)^{-1}$. The efficiency of the estimates (1·6) is then naturally defined as the ratio of the determinants of the covariance matrices $\sigma^2(X'\alpha^{-1}X)^{-1}$ and (1·7). Thus

$$\text{Eff. (b)} = \frac{|X'X|^2}{|X'\alpha X||X'\alpha^{-1}X|}. \tag{3·1}$$

It should be noted that Eff. (b) is invariant when X is transformed to P by $P = XG$. It is always possible to find G so that $P'P = I$. Thus it may be assumed without loss of generality that X satisfies (2·2).

If all the regression vectors are latent vectors† of α, Eff. (b) = 1. This may be verified immediately since only the diagonal terms in $X'X$, $X'\alpha X$ and $X'\alpha^{-1}X$ are then non-zero. The result finds an application in the work of R. L. & T. W. Anderson (1950) on testing for circular serial correlation when a short Fourier series has been fitted by least squares. In this problem the regression vectors are the latent vectors of the quadratic form occurring in the joint density function of successive observations from a circular autoregressive process so that the estimates of the regression coefficients are optimal whether serial correlation, of the type envisaged, is present or not.‡

The expression (3·1) tends of course to unity as $\alpha \to I$. The important question here is how small Eff. (b) may become, where α is fixed and not equal to I but when X is unrestricted. This lower bound to the efficiency may be obtained by generalizing an inequality due§ to J. W. S. Cassels. Cassels's inequality states that for fixed $a_j > 0$, $b_j > 0$ and all $w_j \geqslant 0$ $(j = 1, ..., N)$

$$\frac{4rR}{(r+R)^2} \leqslant \frac{\left(\sum_{1}^{N} a_j b_j w_j\right)^2}{\left(\sum_{1}^{N} a_j^2 w_j\right)\left(\sum_{1}^{N} b_j^2 w_j\right)} \leqslant 1, \tag{3·2}$$

where $r = \min_j a_j/b_j$, $R = \max_j a_j/b_j$. If the numbering of the a_j and b_j is so arranged that $r = a_1/b_1$ and $R = a_N/b_N$, then the lower bound is attained for $w_1 = 1/a_1 b_1$ and $w_N = 1/a_N b_N$, $w_j = 0$ $(j \neq 0, N)$. The lower bound in (3·2) is the square of the ratio of the geometric to the arithmetic mean of r and R.

To obtain the minimum of Eff. (b) by varying X subject to the restriction that the regression vectors are orthogonal, we first eliminate from the problem any regression

---

† In contrast to this, the bias matrix (2·1) is not zero when the regression vectors are latent vectors of α.

‡ This result is, however, little more than a mathematical curiosity. For apart from the fact that the Andersons' test has power against other alternatives, such estimates while optimal cannot be validly tested in the usual way.

§ A proof of this inequality is given in the Appendix. The inequality and the proof were communicated privately by Dr Cassels to the author, who wishes to record his gratitude.

vectors which are latent vectors of α. Suppose that $x_{k-h+1}, \ldots x_k$ are latent vectors of α associated with the latent roots $\alpha_{N-h+1}, \ldots, \alpha_N$. If the remaining x_i have co-ordinates Z_{il} ($l = 1, \ldots, N$) when referred to the latent vectors of α as axes

$$\text{Eff. (b)} = \frac{\prod_{i=1}^{k-h} \left(\sum_{l=1}^{N-h} Z_{il}^2 \right)^2}{\sum_{l=1}^{N-h} \alpha_l Z_{il} Z_{jl} \sum_{l=1}^{N-h} \frac{1}{\alpha_l} Z_{il} Z_{jl}}. \tag{3.3}$$

The matrices $X'\alpha X$ and $X'\alpha^{-1} X$ of (3·1), and therefore the matrices in the denominator of (3·3), are positive-definite by their form. By a theorem of Hadamard (see Hardy, Littlewood & Polya, 1934, p. 34) if c_{ij} are the elements of positive-definite matrix

$$|c_{ij}| \leqslant c_{11} c_{22} \ldots c_{NN}. \tag{3.4}$$

Thus

$$\text{Eff. (b)} \geqslant \prod_{i=1}^{k-h} \frac{\left(\sum_{l=1}^{N-h} Z_{il}^2 \right)^2}{\left(\sum_{l=1}^{N-h} \alpha_l Z_{il}^2 \right) \left(\sum_{l=1}^{N-h} \frac{1}{\alpha_l} Z_{il}^2 \right)}.$$

Without loss of generality we may assume that $\alpha_1 \leqslant \alpha_2 \leqslant \ldots \leqslant \alpha_{N-h}$. We now choose the Z vectors to minimize the right-hand side of (3·4) applying Cassels's inequality successively. When this is done, the matrix X takes the form

$$\left[\frac{a_1 + a_{N-h}}{\sqrt{2}}, \frac{a_2 + a_{N-h-1}}{\sqrt{2}}, \ldots, \frac{a_{k-h} + a_{N-k+1}}{\sqrt{2}}, a_{N-h+1}, \ldots, a_N \right].$$

Using (3·2) with $a_l = \sqrt{\alpha_l}, a_l b_l = 1$ and $w_{il} = Z_{il}^2$, we have the result†

$$\frac{4\alpha_1 \alpha_{N-h}}{(\alpha_1 + \alpha_{N-h})^2} \frac{4\alpha_2 \alpha_{N-h-1}}{(\alpha_2 + \alpha_{N-h-1})^2} \cdots \frac{4\alpha_{k-h} \alpha_{N-k-1}}{(\alpha_{k-h} + \alpha_{N-k+1})^2} \leqslant \text{Eff. (b)} \leqslant 1. \tag{3.5}$$

Thus in spite of the use of Hadamard's inequality, we have arrived at a lower bound for Eff. (b) which is attainable. This is so because we were led to a choice of X for which $X'X$, $X'\alpha X$ and $X'\alpha^{-1} X$ are diagonal. The general rule for the formation of the lower bound to the efficiency is evident. From the $N - h$ roots of α not associated with any regression vectors, we choose successively the $(k-h)$ most extreme pairs. The lower bound to the efficiency thus depends on the ratios of the extreme roots of α. When the roots of α are all approximately equal (approximately equal to unity if the error process is stationary) the efficiency must be high. It will be shown in Part II that high serial correlation of the autoregressive and moving average types leads to α having roots of very different magnitudes and therefore to the possibility of low efficiencies. Finally, it will be noted that the maximum and minimum efficiencies are taken when the arbitrary regression vectors are respectively latent vectors of α and sums of certain pairs of latent vectors of α; using the invariance properties of (3·1) these statements serve to define the maximizing and minimizing regression spaces.

4. BOUNDS ON THE SIGNIFICANCE POINTS OF THE t- AND F-TESTS

The significance tests usually made in regression analysis relate to linear hypotheses about the regression coefficients. Thus in the model (1·1), these hypotheses will be special cases of a null hypothesis specifying the value of one or more linear functions of $\beta_1, \ldots, \beta_k$. When the

† This form of the lower bound was conjectured by J. Durbin.

errors are independently distributed with the same normal distribution, it is well known that the optimal test statistics for such hypotheses have the t- or F-distributions. In this section bounds will be obtained for the distributions of these statistics when the covariance matrix of the errors is $\sigma^2\alpha$; these bounding distributions converge to the tabulated distributions as $\alpha \to I$. We begin by showing that there is no loss of generality in considering only hypotheses specifying the value of several regression coefficients and in assuming that the regression vectors form an orthonormal set. The bounding distributions for this reduced problem will then be derived. The section concludes with a short discussion of the numerical determination of the bounding significance points.

We consider again the model (1·1), $y = X\beta + u$, where the errors u have a multivariate-normal distribution with zero mean vector and covariance matrix $\sigma^2\alpha$ and where the regression vectors are linearly independent. We suppose that the null hypothesis to be tested is

$$f_i'\beta = \phi_i \quad (i = 1, ..., h \leqslant k), \tag{4·1}$$

where $f_i, \phi_i \ (i = 1, ..., h)$ are given and $f_i \ (i = 1, ..., h)$ are linearly independent. The optimal test of the hypothesis (4·1) when $\alpha = I$ is obtained by building up a quadratic form in the variables $f_i'b - \phi_i$ (b is the least-squares estimate of β defined in (1·6)) which is distributed as a multiple of χ^2 with h degrees of freedom. The form must therefore be based on the inverse of the covariance matrix of the variables $f_i'b - \phi_i$. Because σ^2 is unknown it is necessary to divide this form by the estimate s^2 of σ^2 given in (1·9) in order to obtain a test statistic. The ratio so obtained has the F-distribution with h and $N - k$ degrees of freedom provided the null hypothesis is true. This is the statistic we must examine when α is not necessarily equal to I.

If X is transformed to Z by the transformation P

$$Z = XP, \tag{4·2}$$

it is always possible to find a non-singular P so that $Z'Z = I$. For P has only to satisfy

$$P'X'XP = I. \tag{4·3}$$

Defining $\delta = P^{-1}\beta$, where P satisfies (4·3), (1·1) becomes

$$y = Z\delta + u. \tag{4·4}$$

The least-squares estimates of β and δ, b and d say, are given by $b = (X'X)^{-1}X'y$, $d = Z'y$. It is easily seen that

$$d = P^{-1}b. \tag{4·5}$$

The denominator of the test statistic is s^2, and it has been seen in the derivation of (1·10) that s^2 is proportional to

$$u'u - u'X(X'X)^{-1}u, \tag{4·6}$$

which becomes, under the transformation (4·2) and (4·3),

$$u'u - u'Z(Z'Z)^{-1}Z'u = u'u - u'ZZ'u. \tag{4·7}$$

Thus (4·6) is invariant under transformations (4·2) and (4·7) is invariant under transformations (4·2) with $PP' = I$. On the null hypothesis, the variables in the numerator are given by

$$f_i'b - \phi_i = f_i'(X'X)^{-1}X'u \quad (i = 1, ..., h). \tag{4·8}$$

Thus for $i = 1, \ldots, h$, $\mathbf{f}'_i \mathbf{b} - \phi_i$ are distributed as $\mathbf{f}'_i \mathbf{b}$ when $\boldsymbol{\beta} = 0$ or as $\mathbf{f}'_i \mathbf{Pd}$ when $\boldsymbol{\delta} = 0$. We may express the numerator quadratic form in terms of $\mathbf{d}$ without first stating it in terms of $\mathbf{b}$. Writing $\mathbf{f}'_i \mathbf{P} = \mathbf{g}'_i$, it must on the null hypothesis have the form

$$[\mathbf{g}'_1 \mathbf{d}, \ldots, \mathbf{g}'_h \mathbf{d}] \begin{bmatrix} \mathbf{g}'_1 \mathbf{g}_1 & \cdots & \mathbf{g}'_1 \mathbf{g}_h \\ \vdots & & \vdots \\ \mathbf{g}'_h \mathbf{g}_1 & \cdots & \mathbf{g}'_h \mathbf{g}_h \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}'_1 \mathbf{d} \\ \vdots \\ \mathbf{g}'_h \mathbf{d} \end{bmatrix}, \tag{4.9}$$

because this is seen to be correct when $\boldsymbol{\alpha} = \mathbf{I}$. (4.9) may be written as

$$\mathbf{d}'[\mathbf{g}_1, \ldots, \mathbf{g}_h] \begin{bmatrix} \mathbf{g}'_1 \mathbf{g}_1 & \cdots & \mathbf{g}'_1 \mathbf{d}_h \\ \vdots & & \vdots \\ \mathbf{g}'_h \mathbf{g}_1 & \cdots & \mathbf{g}'_h \mathbf{g}_h \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}'_1 \\ \vdots \\ \mathbf{g}'_h \end{bmatrix} \mathbf{d}. \tag{4.10}$$

The matrix of this quadratic form in $\mathbf{d}$ is symmetric and has h unit roots and $k - h$ zero roots. Thus an orthogonal transformation $\mathbf{H}$ may be found so that if $\mathbf{d} = \mathbf{Hc}$, (4.10) reduces to $\mathbf{c}'\mathbf{Jc}$, where

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_h & \mathbf{0}_1 \\ \mathbf{0}_2 & \mathbf{0}_3 \end{bmatrix},$$

a $k \times k$ matrix with units in the first h places on the leading diagonal and zeros elsewhere. Since $\mathbf{H}$ is orthogonal,

$$\mathbf{c} = \mathbf{H}^{-1}\mathbf{d} = \mathbf{H}'\mathbf{Z}'\mathbf{u}. \tag{4.11}$$

If finally we put

$$\mathbf{W} = \mathbf{ZH}, \tag{4.12}$$

then

$$\mathbf{c} = \mathbf{W}'\mathbf{u},$$

and the test statistic is proportional to

$$\frac{\mathbf{u}'\mathbf{WJW}'\mathbf{u}}{\mathbf{u}'\mathbf{u} - \mathbf{u}'\mathbf{WW}'\mathbf{u}} = \frac{(\mathbf{u}'\mathbf{w}_1)^2 + \ldots + (\mathbf{u}'\mathbf{w}_h)^2}{\mathbf{u}'\mathbf{u} - (\mathbf{u}'\mathbf{w}_1)^2 - \ldots - (\mathbf{u}'\mathbf{w}_h)^2}, \tag{4.13}$$

where $\mathbf{w}_1, \ldots, \mathbf{w}_k$ form an orthonormal set of vectors. This is the desired canonical form for the test statistic because it is the form which arises in testing the hypothesis

$$\psi_i = \psi_i^0 \quad (i = 1, \ldots, h)$$

in the model

$$\mathbf{y} = \mathbf{W}\boldsymbol{\psi} + \mathbf{u},$$

where $\mathbf{W}'\mathbf{W} = \mathbf{I}$ and $\boldsymbol{\psi} = \{\psi_1, \ldots, \psi_k\}$. $\mathbf{c}$ defined above is the least-squares estimate of $\boldsymbol{\psi}$.

Since the proof above for the general linear hypothesis presents some difficulty, it seems worth while to establish the result for a simple particular case. Suppose that the null hypothesis is

$$\boldsymbol{\beta} = \boldsymbol{\beta}_0. \tag{4.14}$$

To obtain the form of the test statistic, we refer to (1.8) which defines the covariance matrix of $\mathbf{b} - \boldsymbol{\beta}_0$ and (1.9) which gives an expression for s^2. Thus the statistic is proportional to

$$\frac{(\mathbf{b} - \boldsymbol{\beta}_0)'(\mathbf{X}'\mathbf{X})(\mathbf{b} - \boldsymbol{\beta}_0)}{(\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb})}. \tag{4.15}$$

Since $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, the statistic (4.15) on the null hypothesis may be put in the form

$$\frac{\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}}{\mathbf{u}'\mathbf{u} - \mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}}. \tag{4.16}$$

If now $\mathbf{X}$ is transformed to $\mathbf{Z}$, so that $\mathbf{Z}'\mathbf{Z} = \mathbf{I}$, the discussion of (4.2) ... (4.7) shows that (4.16) is reduced to

$$\frac{\mathbf{u}'\mathbf{ZZ}'\mathbf{u}}{\mathbf{u}'\mathbf{u} - \mathbf{u}'\mathbf{ZZ}'\mathbf{u}}. \tag{4.17}$$

This is the required canonical form for the hypothesis (4.14).

We consider the model of (1·1) and (1·2) under the assumptions of orthonormality (2·2). The errors u will now be assumed to have a multivariate normal distribution with zero mean vector and covariance matrix $\sigma^2\alpha$. We suppose initially that the first $k-1$ of the regression vectors x are latent vectors of α, the remaining vector being arbitrary. Denote $x_i = a_i$ where $\alpha a_i = \alpha_i a_i$ $(i = 1, \ldots, k-1)$. To test the hypothesis

$$\beta_k = \beta_k^0 \tag{4·18}$$

the appropriate statistic is

$$t^2 = \frac{(N-k)(b_k - \beta_k^0)^2}{\sum\limits_{i=1}^{N}\left(y_i - \sum\limits_{1}^{k-1} b_i a_{ii} - b_k x_{ki}\right)^2}. \tag{4·19}$$

On the null hypothesis $t^2/(N-k)$ has the matrix form

$$\frac{t^2}{N-k} = \frac{u'x_k x_k' u}{u'(I - a_1 a_1' - \ldots - a_{k-1}a_{k-1}' - x_k x_k')u} = T, \quad \text{say.} \tag{4·20}$$

Bounds on the significance points of T will now be found which enclose the true significance point of T, no matter what the direction of the vector x_k.

By a device, often used in the treatment of ratios, we have

$$P(T < \tau) = P(u'(x_k x_k' - \tau(I - a_1 a_1' - \ldots - a_{k-1}a_{k-1}' - x_k x_k'))u < 0). \tag{4·21}$$

But the quadratic form on the right-hand side of (4·21) is distributed like $\sum\limits_{i=1}^{N} \nu_i v_i^2$, where ν_i are the roots of the determinantal equation

$$|x_k x_k' - \tau(I - a_1 a_1' - \ldots - a_{k-1}a_{k-1}' - x_k x_k') - \nu\alpha^{-1}| = 0, \tag{4·22}$$

and the v_i are N.I.D. $(0, 1)$. For fixed τ, let x_k vary. Then the ν_i and $P(\sum \nu_i v_i^2 < 0)$ vary. We will show that there exists a direction of x_k such that *all* the ν_i take maximum values. For this x_k, $P(\sum \nu_i v_i^2 < 0)$ is clearly a minimum. Since, for fixed x_k, $P(T < \tau)$ is a monotonically increasing function of τ, it follows that the significance point of T, τ_i, say, for this special maximizing direction is greater than the significance point at the same level for any other vector x_k. In the same way a lower bound τ_L will be established.

For this purpose apply to (4·22) the orthogonal transformation K which diagonalizes α. If

$$K'x_k = m,$$

$$K'\alpha^{-1}K = \begin{bmatrix} 1/\alpha_1 & & & \\ & 1/\alpha_2 & & \\ & & \ddots & \\ & & & 1/\alpha_N \end{bmatrix}, \tag{4·23}$$

then the first $k-1$ co-ordinates of m are zero because x_k is required to be orthogonal to the $k-1$ latent vectors corresponding to $\alpha_1, \ldots, \alpha_{k-1}$, and

$$K'(I - a_1 a_1' - \ldots - a_{k-1}a_{k-1}')K = \begin{bmatrix} 0 & & & & \\ & \cdots & & & \\ & & 0 & & \\ & & & 1 & \\ & & & & \cdots \\ & & & & & 1 \end{bmatrix}, \tag{4·24}$$

where the matrix on the right-hand side of (4·24) is an $N \times N$ identity matrix with the first $k-1$ diagonal elements replaced by zeros. Thus equation (4·22) has $k-1$ zero roots and $N-k+1$ which satisfy the reduced determinantal equation

$$
\begin{vmatrix}
(1+\tau)m_k^2 - \tau - \dfrac{\nu}{\alpha_k}, & (1+\tau)m_k m_{k+1}, & \cdots, & (1+\tau)m_k m_N \\[2mm]
(1+\tau)m_{k+1}m_k, & (1+\tau)m_{k+1}^2 - \tau - \dfrac{\nu}{\alpha_{k+1}}, & \cdots, & (1+\tau)m_{k+1}m_N \\
\vdots & \vdots & & \vdots \\
(1+\tau)m_N m_k, & (1+\tau)m_N m_{k+1}, & \cdots, & (1+\tau)m_N^2 - \tau\dfrac{\nu}{\alpha_N}
\end{vmatrix} = 0. \quad (4\cdot25)
$$

Subtracting $m_{k+1}\, m_k$ times the first from the second row, etc., and expanding the result as a bordered determinant we find the equation in ν.

$$
\prod_{p=k}^{N}\left(\tau + \frac{\nu}{\alpha_p}\right) - (1+\tau)\sum_{p=k}^{N} m_p^2 \prod_{\substack{j=k \\ j\neq p}}^{N}\left(\tau + \frac{\nu}{\alpha_j}\right) = 0. \quad (4\cdot26)
$$

i.e.

$$
\sum_{k}^{N} \frac{m_p^2}{\tau + \dfrac{\nu}{\alpha_p}} = \frac{1}{1+\tau}. \quad (4\cdot27)
$$

We see that the effect of the regression variables chosen as latent vectors has been the same as on previous occasions, i.e. the corresponding latent roots are eliminated from the problem. We may therefore without loss of generality suppose that

$$
\alpha_k < \alpha_{k+1} < \dots < \alpha_N. \quad (4\cdot28)
$$

The location of the roots ν_i is easily seen graphically from (4·27). We may, however, proceed as follows. Define the continuous function of ν,

$$
f(\nu) = \prod_{p=k}^{N}\left(\tau + \frac{\nu}{\alpha_p}\right) - (1+\tau)\sum_{p=k}^{N} m_p^2 \prod_{\substack{j=k \\ j\neq p}}^{N}\left(\tau + \frac{\nu}{\alpha_j}\right).
$$

Then since

$$
f(-\tau\alpha_i) = -(1+\tau)m_i^2 \tau^{N-k}\prod_{\substack{j=k \\ j\neq i}}^{N}\left(1 - \frac{\alpha_i}{\alpha_j}\right)
$$

is seen to have the opposite sign to $f(-\tau\alpha_{i+1})$, the equation (4·26) must have at least one root in each of the intervals $(-\tau\alpha_{i+1}, -\tau\alpha_i)$.

But, considering equation (4·27), we see that

$$
\sum_{k}^{N} \frac{m_i^2}{\tau + \dfrac{\alpha_N}{\alpha_i}} \leqslant \frac{1}{1+\tau}, \quad \sum_{k}^{N} \frac{m_i^2}{\tau + \dfrac{\alpha_k}{\alpha_i}} \geqslant \frac{1}{1+\tau}
$$

by (4·28) and the fact that $\sum_{k}^{N} m_i^2 = 1$. Hence there is at least one root of (4·26) and (4·27) in the interval (α_k, α_N).

Combining the results, we have proved that there is one root ν_i in each of the $N-k+1$ intervals

$$
(-\tau\alpha_{i+1}, -\tau\alpha_i) \quad (i = k, \dots, N-1), \qquad (\alpha_k, \alpha_N). \quad (4\cdot29)
$$

Finally, if x_k is chosen to be the latent vector associated with the root α_k, i.e. a_k, (4·26) gives us the roots

$$-\tau\alpha_N, \quad \ldots, \quad -\tau\alpha_{k+1} \quad \text{and} \quad \alpha_k, \tag{4·30}$$

which are seen from (4·29) to be the least possible set of roots. If x_k is chosen to be the latent vector associated with α_N, i.e. a_N, (4·26) gives the greatest possible set of roots

$$-\tau\alpha_{N-1}, \quad \ldots, \quad -\tau\alpha_k \quad \text{and} \quad \alpha_N. \tag{4·31}$$

We have therefore proved

THEOREM 1. *If τ is such that $P(T < \tau) = C$ $(0 \leqslant C \leqslant 1)$, T being defined by (4·20), then there exist numbers τ_L, τ_U such that*

$$\tau_L \leqslant \tau \leqslant \tau_U \tag{4·32}$$

for all vectors x_k. τ_L and τ_U are determined by the equations

$$P\left(\frac{\alpha_k \eta_k^2}{\alpha_{k+1}\eta_{k+1}^2 + \ldots + \alpha_N \eta_N^2} < \tau_L\right) = C, \left.\vphantom{\frac{\alpha_k}{\alpha}}\right\}$$
$$P\left(\frac{\alpha_N \eta_N^2}{\alpha_k \eta_k^2 + \ldots + \alpha_{N-1}\eta_{N-1}^2} < \tau_U\right) = C, \left.\vphantom{\frac{\alpha_k}{\alpha}}\right\} \tag{4·33}$$

where $\eta_k, \ldots, \eta_{N-1}$ are N.I.D. $(0, 1)$.

We now extend this theorem, with the same assumptions about the model, to a joint test of the null hypothesis

$$\beta_1 = \beta_1^0, \quad \ldots, \quad \beta_h = \beta_h^0 \quad (h \leqslant k-1), \qquad \beta_k = \beta_k^0. \tag{4·34}$$

The statistic T for this hypothesis is on the null hypothesis

$$T = \frac{u'a_1 a_1' u + \ldots + u'a_h a_h' u + u'x_k x_k' u}{u'(I - a_1 a_1' - \ldots - a_{k-1} a_{k-1}' - x_k x_k') u}. \tag{4·35}$$

Proceeding as before the determinantal equation is found to be

$$| a_1 a_1' + \ldots + a_h a_h' + x_k x_k' - \tau(I - a_1 a_1' - \ldots - a_{k-1} a_{k-1}' - x_k x_k') - \nu\alpha^{-1} | = 0. \tag{4·36}$$

After applying the orthogonal transformation K of (4·23) the roots ν_i are seen to be

$$\alpha_1, \alpha_2, \ldots, \alpha_h, \quad k-h-1 \text{ zeros,}$$

and the $N - k + 1$ roots of (4·25). This leads immediately to the following extension of Theorem 1.

THEOREM 2. *If τ is such that $P(T < \tau) = C$, $(0 \leqslant C \leqslant 1)$ T being defined by (4·35), then there exist numbers τ_L and τ_U such that*

$$\tau_L \leqslant \tau \leqslant \tau_U$$

for all vectors x_k. τ_L and τ_U are determined by the equations

$$P\left(\frac{\alpha_1 \eta_1^2 + \ldots + \alpha_h \eta_h^2 + \alpha_k \eta_k^2}{\alpha_{k+1}\eta_{k+1}^2 + \ldots + \alpha_N \eta_N^2} < \tau_L\right) = C, \left.\vphantom{\frac{\alpha_1}{\alpha}}\right\}$$
$$P\left(\frac{\alpha_1 \eta_1^2 + \ldots + \alpha_h \eta_h^2 + \alpha_N \eta_N^2}{\alpha_k \eta_k^2 + \ldots + \alpha_{N-1}\eta_{N-1}^2} < \tau_U\right) = C, \left.\vphantom{\frac{\alpha_1}{\alpha}}\right\} \tag{4·37}$$

where $\eta_1, \ldots, \eta_h, \eta_k, \ldots, \eta_N$ are N.I.D. $(0, 1)$.

To complete this case, we examine the test of the null hypothesis

$$\beta_1 = \beta_1^0, \ldots, \beta_h = \beta_h^0 \quad (h \leqslant k-1). \tag{4.38}$$

The T statistic is here defined on the null hypothesis by

$$T = \frac{u'a_1 a_1' u + \ldots + u'a_h a_h' u}{u'(I - a_1 a_1' - \ldots - a_{k-1} a_{k-1}' - x_k x_k') u}. \tag{4.39}$$

The associated determinantal equation is

$$| a_1 a_1' + \ldots + a_h a_h' - \tau(I - a_1 a_1' - \ldots - a_{k-1} a_{k-1}' - x_k x_k') - \nu\alpha^{-1} | = 0. \tag{4.40}$$

The roots of (4.40) are $\alpha_1, \ldots, \alpha_h$, $k-h$ zeros and the roots of the determinantal equation

$$
\begin{vmatrix}
\tau m_k^2 - \tau - \dfrac{\nu}{\alpha_k}, & \tau m_k m_{k+1}, & \ldots, & \tau m_k m_N \\[2mm]
\tau m_{k+1} m_k, & \tau m_{k+1}^2 - \tau - \dfrac{\nu}{\alpha_{k+1}}, & \ldots, & \tau m_{k+1} m_N \\
\vdots & \vdots & & \vdots \\
\tau m_N m_k, & \tau m_N m_{k+1}, & \ldots, & \tau m_N^2 - \tau - \dfrac{\nu}{\alpha_N}
\end{vmatrix}
= 0. \tag{4.41}
$$

Expanding as before, we find the equation in ν

$$\prod_{i=k}^{N} \left(\tau + \frac{\nu}{\alpha_i} \right) - \tau \sum_{i=k}^{N} m_i^2 \prod_{j \neq i} \left(\tau + \frac{\nu}{\alpha_j} \right) = 0, \tag{4.42}$$

i.e.

$$\frac{1}{\tau} = \sum_{i=k}^{N} \frac{m_i^2}{\tau + \dfrac{\nu}{\alpha_i}}. \tag{4.43}$$

As with equation (4.26), it can be shown that equation (4.42) has at least one root in each of the intervals

$$(-\tau\alpha_{i+1}, -\tau\alpha_i).$$

By inspection we see that it always has a zero root so that there is just one in each of these intervals. If x_k is chosen to be the latent vector of α associated with the root α_k, i.e. a_k, the least set of roots for ν_i is obtained, while if x_k is the latent vector of α associated with the root α_N, i.e. a_N, the greatest set is found. We may thus state

THEOREM 3. If τ is such that $P(T < \tau) = C \, (0 \leqslant C \leqslant 1)$, T being defined by (4.39), then there exist numbers τ_L and τ_U such that $\tau_L \leqslant \tau \leqslant \tau_U$

for all vectors x_k. τ_L and τ_U are determined by the equations

$$
P\left(\frac{\alpha_1 \eta_1^2 + \ldots + \alpha_h \eta_h^2}{\alpha_{k+1} \eta_{k+1}^2 + \ldots + \alpha_N \eta_N^2} < \tau_L \right) = C,
$$
$$
P\left(\frac{\alpha_1 \eta_1^2 + \ldots + \alpha_h \eta_h^2}{\alpha_k \eta_k^2 + \ldots + \alpha_{N-1} \eta_{N-1}^2} < \tau_U \right) = C,
\tag{4.44}
$$

where $\eta_1 \ldots \eta_h, \eta_k \ldots \eta_N$ are N.I.D. $(0, 1)$.

We now proceed to a case where there is more than one arbitrary regression vector. Suppose that all k regression vectors are arbitrary except for restriction (2.2) and that the hypothesis to be tested is

$$\beta_i = \beta_i^0 \quad (i = 1, \ldots, k). \tag{4.45}$$

On the null hypothesis the usual F-ratio is distributed like a multiple of

$$T = \frac{u'XX'u}{u'(I-XX')u};$$ (4·46)

therefore

$$R = \frac{T}{1+T} = \frac{u'XX'u}{u'u}.$$ (4·47)

Since T is a monotonic function of $R \left(T = \frac{R}{1-R} \right)$, the extremes of $P(R < r)$ will yield extremes of $P(T < \tau)$. We therefore proceed to find distributions which bound that of R above and below. If H is the linear transformation such that

$$H\alpha H' = I,$$ (4·48)

then the variables $v = Hu$ are N.I.D. $(0, 1)$—for σ^2 may be set equal to unity without loss of generality. Then

$$R = \frac{v'H'^{-1}XX'H^{-1}v}{v'(HH')^{-1}v}.$$

From (4·48),

$$(HH')^{-1} = \begin{bmatrix} \alpha_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_N \end{bmatrix},$$

so that we may write

$$R = \frac{v'H'^{-1}XX'H^{-1}v}{\sum\limits_{1}^{N} \alpha_i v_i^2}.$$ (4·49)

R is a ratio of non-negative definite quadratic forms in N.I.D. $(0, 1)$ variables. For variation of X, the upper and lower bounding random variables for R are obtained when the numerator takes its least and greatest forms. Now the numerator of R is distributed like

$$\sum_{i=1}^{k} \nu_i \eta_i^2,$$ (4·50)

where $\eta_1, ..., \eta_k$ are N.I.D. $(0, 1)$ and $\nu_1, ..., \nu_k$ are the k non-zero roots, written in ascending order of magnitude, of

$$| H'^{-1}XX'H^{-1} - \nu I | = 0,$$ (4·51)

i.e. the non-zero roots of

$$| H^{-1}H'^{-1}XX' - \nu I | = 0,$$

since the roots of a product are indifferent to the order of multiplication. But by (4·48)

$$H^{-1}H'^{-1} = (H'H)^{-1} = \alpha,$$

so that $\nu_1 < ... < \nu_k$ are the k non-zero roots of

$$| \alpha XX' - \nu I | = 0.$$ (4·52)

Suppose that $z_1, ..., z_{N-k}$ are vectors which together with $x_1, ..., x_k$ form a complete orthonormal set in N-space. If $Z = [z_1, ..., z_{N-k}]$, then,

$$[X, Z][X, Z]' = [X, Z]'[X, Z]$$
$$= I.$$

But
$$[X, Z][X, Z]' = XX' + ZZ',$$

so that
$$XX' = I - ZZ'$$
$$= (I - z_1 z_1') \dots (I - z_{N-k} z_{N-k}')$$
$$= M_1 \dots M_{N-k}, \quad \text{say}.$$

where $M_1, \dots, M_{N-k}$ are idempotent matrices of rank $N-1$.

With this characterization of XX', inequalities on the roots of (4·52) may be obtained immediately from the investigation of Durbin & Watson (1950). They show, with a proof similar to that of Theorem 1, that the roots of αM_1 lie between the roots of α. Thus the non-zero roots of $\alpha M_1 M_2$ lie between the non-zero roots of αM_1. Combining the inequalities given by successive applications of this result, it follows that the roots $\nu_1, \dots, \nu_k$ of (4·52) must satisfy the inequalities

$$\alpha_1 \leqslant \nu_1 \leqslant \alpha_{N-k+1}, \quad \alpha_2 \leqslant \nu_2 \leqslant \alpha_{N-k+2}, \quad \dots, \quad \alpha_k \leqslant \nu_k \leqslant \alpha_N. \tag{4·53}$$

The least set of roots is obtained when the k regression vectors are chosen to be the latent vectors $a_1, \dots, a_k$ (or linear combinations of these vectors). The greatest set of roots is obtained when the regression vectors are similarly related to the latent vectors $a_{N-k+1}, \dots, a_N$. These choices lead to the extreme forms of R required. We may thus state

THEOREM 4. If τ is such that $P(T < \tau) = C \ (0 \leqslant C \leqslant 1)$, T being defined by (4·46), then there exist numbers τ_L and τ_U such that
$$\tau_L \leqslant \tau \leqslant \tau_U$$

for all X. τ_L and τ_U are defined by the equations

$$P\left(\frac{\alpha_1 \eta_1^2 + \dots + \alpha_k \eta_k^2}{\alpha_{k+1} \eta_{k+1}^2 + \dots + \alpha_N \eta_N^2} < \tau_L\right) = C,$$
$$P\left(\frac{\alpha_{N-k-1} \eta_{N-k+1}^2 + \dots + \alpha_N \eta_N^2}{\alpha_1 \eta_1^2 + \dots + \alpha_{N-k} \eta_{N-k}^2} < \tau_U\right) = C. \tag{4·54}$$

where $\eta_1, \dots, \eta_N$ are N.I.D. $(0, 1)$.

The rules of formation of the random variables which lead to the bounding significance points are now evident from Theorems 1, 2, 3, 4 so that there is no need to give the details for every type of test which may arise. In all cases it is seen that the distributions of the bounding variables converge to the appropriate F-distribution as the α's approach equality.

The calculation of the bounding significance points discussed above presents some difficulty. We are faced with the determination of the significance points of ratio of two independent weighted sums of squares of normal deviates. Since the weights are all positive, the method of Pitman & Robbins (1949) may be used when the α's are not too different. Some numerical results obtained by this method will be given in Part II. The amount of arithmetic required is always large—the electronic computer of the Cambridge Mathematical Laboratory was used in these calculations—so it is worth while examining approximate methods.

To illustrate the approximations used in Part II, let us consider the random variable which generates the lower bound in Theorem 4, namely,

$$\frac{\alpha_1 \eta_1^2 + \dots + \alpha_k \eta_k^2}{\alpha_{k+1} \eta_{k+1}^2 + \dots + \alpha_N \eta_N^2}.$$

The simplest approximation to this ratio is

$$\frac{\sum\limits_{1}^{k} x_i}{k} \cdot \frac{N-k}{\sum\limits_{k+1}^{N} \alpha_i} \cdot \frac{\eta_1^2 + \ldots + \eta_k^2}{\eta_{k+1}^2 + \ldots + \eta_N^2}, \quad \text{i.e.} \quad \frac{\sum\limits_{1}^{k} x_i}{\sum\limits_{k+1}^{N} x_i} F_{k,N-k},$$

where $F_{k,N-k}$ has the F-distribution with k and $N-k$ degrees of freedom. If in a practical analysis we had some knowledge of the correlogram of the errors, this approach suggests that multiples of the usual significance points might be used to make a reasonable test. A better approximation may be obtained by replacing the numerator and denominator by multiples of σ^2 with the correct variances as well as the correct means. The significance points of this approximation must be obtained by interpolation and do not have the simple interpretation of the first approximation.

The author has pleasure in recording his indebtedness to Mr J. Durbin for many helpful discussions.

APPENDIX

A proof of the inequality (3·2)

THEOREM. Let $a_j > 0$, $b_j > 0$, $w_j \geqslant 0$ $(j = 1, \ldots, n)$. Then

$$1 \leqslant \frac{(\Sigma a_j^2 w_j)(\Sigma b_j^2 w_j)}{(\Sigma a_j b_j w_j)^2} \leqslant \max_{i,j} \frac{(a_i b_j + a_j b_i)^2}{4 a_i a_j b_i b_j}.$$

The extreme values are attained where at most two of $w_1, \ldots, w_n$ are non-zero. This theorem and its proof below are due to J. W. S. Cassels.

Proof. As a straightforward extremal problem it may be shown that

$$1 \leqslant \frac{(1+kw)(1+k^{-1}w)}{(1+w)^2} \leqslant \frac{(1+k)(1+k^{-1})}{4} \quad \text{if} \quad k > 0, w \geqslant 0.$$

Thus the theorem holds for $n = 2$—for put

$$w = \frac{a_2 b_2 w_2}{a_1 b_1 w_1}, \quad k = \frac{a_1 b_2}{a_2 b_1}.$$

Furthermore, the extremes of the theorem are attained when at most two of $w_1, \ldots, w_n$ are non-zero. It may be assumed without loss of generality that

$$\begin{vmatrix} a_i^2 & a_j^2 & a_k^2 \\ b_i & b_j & b_k \\ a_i b_i & a_j b_j & a_k b_k \end{vmatrix} \neq 0.$$

For this determinant is given by $\prod\limits_{i,j,k} \begin{vmatrix} a_i & a_j \\ b_i & b_j \end{vmatrix}$, so that if it vanishes we have say

$$a_i = \lambda a_j, \quad b_i = \lambda b_j, \quad \text{where} \quad \lambda > 0.$$

But if this is so the problem is reduced to one in $n-1$ variables.

To prove that the extrema of $\Sigma a_j^2 w_j / (\Sigma b_j^2 w_j + \Sigma a_j b_j w_j)^2$ are attained when at most two of $w_j \ldots w_N$ are non-zero let M be an extreme which is attained when say $w_1 \neq 0, w_2 \neq 0, w_3 \neq 0$. Then we have

$$\frac{d}{dw_k}\{(\Sigma a_j^2 w_j)(\Sigma b_j^2 w_j) - M(\Sigma a_j b_j w_j)^2\} = 0 \quad (k = 1, 2, 3),$$

and so

$$a_k^2 X + b_k^2 Y - 2Ma_k b_k Z = 0 \quad (k = 1, 2, 3),$$

where $X = \Sigma a_j^2 w_j$, $Y = \Sigma b_j^2 w_j$, $Z = \Sigma a_j b_j w_j$. But the determinant of these equations does not vanish. Hence

$$X = Y = MZ = 0,$$

which is impossible.

This establishes the theorem. To obtain the inequality (3·2) it is only necessary to notice that the upper bound in theorem may be put in the form

$$\max_{i,j} \frac{1}{4}\left(\frac{r_i}{r_j} + \frac{r_j}{r_i}\right)^2,$$

where $r_i = a_i/b_i$. Since the function $w + 1/w$ takes its maximum at the end of the range of variation of w, (3·2) follows.

REFERENCES

AITKEN, A. C. (1934-5). On least squares and linear combinations of observations. *Proc. Roy. Soc. Edinb.* **55**, 42–8.

AITKEN, A. C. (1948). On the estimation of many statistical parameters. *Proc. Roy. Soc. Edinb.* **62**, 369.

ANDERSON, R. L. & ANDERSON, T. W. (1950). Distribution of the circular serial correlation coefficient for residuals from a fitted Fourier series. *Ann. Math. Statist.* **21**, 59–81.

BOX, G. E. P. (1954). Some theorems on quadratic forms applied to the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two way classification. *Ann. Math. Statist.* **25**, 484–98.

COCHRANE, D. & ORCUTT, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *J. Amer. Statist. Ass.* **44**, 32–61.

DURBIN, J. & WATSON, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika*, **37**, 409–28.

DURBIN, J. & WATSON, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika*, **38**, 159–78.

HANNAN, E. J. (1955). Exact tests for serial correlation. *Biometrika*, **42**, 133.

HARDY, G. H., LITTLEWOOD, J. E. & POLYA, G. (1934). *Inequalities*. Cambridge University Press.

PITMAN, E. J. G. & ROBBINS, H. (1949). Application of the method of mixtures to quadratic forms in normal variates. *Ann. Math. Statist.* **20**, 552–60.

WOLD, H. (1949). On least squares regression with autocorrelated variables and residuals. *Trans. Int. Inst. Statist.* Reprint, pp. 1–13.

SOME THEOREMS AND SUFFICIENCY CONDITIONS FOR THE MAXIMUM-LIKELIHOOD ESTIMATOR OF AN UNKNOWN PARAMETER IN A SIMPLE MARKOV CHAIN

BY J. GANI

Australian National University, Canberra, A.C.T.

I. SUMMARY

The paper begins with proofs of the usual theorems for the optimum properties of the maximum-likelihood estimator of an unknown parameter θ which defines the transition probabilities $p_{ij}(\theta)$ of a simple ergodic Markov chain. By an ergodic chain is meant one for which, not only is the final chain stationary, but also all possible initial states remain permanently available; these conditions are sufficient to prove that the maximum-likelihood estimator is consistent, and asymptotically normally distributed.

The paper proceeds to establish the form of the transition probabilities $p_{ij}(\theta)$ which admit a sufficient estimator of θ. To do this, the form of the likelihood function admitting a sufficient estimator when the parent distribution is discrete is first derived: this is used to obtain the form of the probabilities $p_i(\theta)$ for a multinomial distribution admitting a sufficient estimator of θ, and the result is finally generalized for the transition probabilities $p_{ij}(\theta)$ of the simple ergodic Markov chain.

The paper closes with an examination of possible forms for the matrix p of transition probabilities $p_{ij}(\theta)$, and these are illustrated with simple examples for Markov chains with two and three states.

II. THE LIKELIHOOD EQUATION FOR A SIMPLE MARKOV CHAIN

Consider the simple Markov chain with s possible states $E_1, \ldots, E_s$, and the matrix of transsition probabilities $p_{ij}(\theta) = \Pr(E_j \mid E_i) \, (i, j = 1, \ldots, s)$, which are all functions of an unknown parameter θ. The matrix is given by

$$
\mathbf{p}(\theta) = \begin{pmatrix} p_{11}(\theta) & \cdots & p_{1s}(\theta) \\ \cdots\cdots\cdots\cdots\cdots \\ p_{s1}(\theta) & \cdots & p_{ss}(\theta) \end{pmatrix},
$$

where the transition probabilities are subject to the condition that

$$
\sum_{j=1}^{s} p_{ij} = 1, \quad \text{for all } i = 1, 2, \ldots, s. \tag{1}
$$

If a realization of the chain results in an observed sequence S of $n+1$ states in the following order

$$
E_i, E_j, \ldots, E_k, E_l,
$$

then it is possible to write the probabilities of this sequence as

$$
\phi(S) = a_i(\theta)\, p_{ij}(\theta) \ldots p_{kl}(\theta), \tag{2}
$$

where $a_i(\theta)$ $(i = 1, \ldots, s)$, the initial probability distribution of the states $E_1, \ldots, E_s$, is assumed known.

The likelihood function will be given by

$$L = \ln \phi(S) = \ln a_i + \ln p_{ij} + \ldots + \ln p_{hi}.$$

Grouping together the transitions from states E_i to states E_j, $(i, j = 1, \ldots, s)$, and denoting the frequency of these transitions by n_{ij}, we write this likelihood function as

$$L = \ln a_i + \sum_{i=1}^{s} \sum_{j=1}^{s} n_{ij} \ln p_{ij},$$

or for simplicity

$$L = \ln a_i + \sum_{i,j} n_{ij} \ln p_{ij}. \tag{3}$$

It follows that, providing $a_i(\theta)$, $p_{ij}(\theta)$ are differentiable with respect to θ, the derivative of the likelihood is

$$\frac{dL}{d\theta} = \frac{1}{a_i}\frac{da_i}{d\theta} + \sum_{i,j} \frac{n_{ij}}{p_{ij}}\frac{dp_{ij}}{d\theta}. \tag{4}$$

As n increases, the second parts of (3) and (4) become dominant, and the likelihood function and its derivative will be

$$L \sim \sum_{i,j} n_{ij} \ln p_{ij}, \tag{5}$$

and

$$\frac{dL}{d\theta} \sim \sum_{i,j} \frac{n_{ij}}{p_{ij}}\frac{dp_{ij}}{d\theta}. \tag{6}$$

The maximum-likelihood estimator T of θ will in this case be given by

$$\left(\frac{dL}{d\theta}\right)_T \sim \left(\sum_{i,j} \frac{n_{ij}}{p_{ij}}\frac{dp_{ij}}{d\theta}\right)_T = 0. \tag{7}$$

It is important to note that the n_{ij} are not linearly independent; the number of transitions from the state E_i to the states $E_1, \ldots, E_s$, will, except for a possible end-effect, be equal to the number of transitions from the states $E_1, \ldots, E_s$, to the state E_i, so that

$$\sum_{j=1}^{s} n_{ij} \doteq \sum_{r=1}^{s} n_{ri}, \tag{8}$$

where the sign $\doteq$ indicates equality or a possible difference of 1 between the sums. As n increases, we may therefore accept the equations

$$\sum_{j=1}^{s} n_{ij} = \sum_{r=1}^{s} n_{ri} = n_i \quad (i = 1, \ldots, s), \tag{9}$$

where

$$\sum_{i=1}^{s} n_i = \sum_{i=1}^{s} \sum_{j=1}^{s} n_{ij} = n.$$

III. OPTIMUM PROPERTIES OF THE MAXIMUM-LIKELIHOOD ESTIMATOR

Following closely the proofs given by Cramér (1946) and Rao (1952) for continuous distributions, we shall deduce that for an ergodic chain, the maximum-likelihood estimator T obtained as a solution of (7) has the following optimum properties:

(a) The estimator T obtained as a solution of (7) is consistent.

(b) The consistent solution of (7) is asymptotically normally distributed about the true value θ_0: it is fully efficient.

A further optimum property due to Huzurbazar (1948) can also be proved; it is that

(c) Any consistent solution of (7) is such that for $n \to \infty$ the probability that the likelihood tend to a maximum converges to 1: the consistent solution is therefore unique.

The proof of this property for the case of the chain considered is, however, identical in its essentials with that given by Huzurbazar for continuous distributions; it will therefore be omitted. In order to establish these properties, we require the assumptions that

$$\frac{dp_{ij}}{d\theta}, \quad \frac{d^2p_{ij}}{d\theta^2}, \quad \frac{d^3p_{ij}}{d\theta^3}$$

exist and are continuous for every θ in a range R including the true value θ_0. Before proceeding to prove these theorems, it may be useful to mention some properties of ergodic chains which we shall refer to.

For an ergodic chain, all possible initial states remain permanently available, and the final chain is stationary and independent of initial conditions. Although it is possible for some transition probabilities p_{ij} of an ergodic chain to be zero, the stationary probabilities $P_i = \Pr(E_i)$ $(i = 1, ..., s)$, which are given by the matrix equation

$$\mathbf{p'P} = \mathbf{P},$$

together with the equation $\sum_i P_i = 1$, where $\mathbf{P}$ is the column vector of elements P_i, will all be non-zero; it is clear that the P_i will also be functions of the parameter θ. In equation (2) no assumption was made about the chain being initially stationary, so that the a_i are not necessarily equal to the P_i; however, if the process is initially stationary, we can write $a_i = P_i$ in this and the subsequent equations (3) and (4). In all that follows, we shall for simplicity consider an initially stationary chain; the results obtained will also hold asymptotically for a chain which is not initially stationary.

The following set of results for an ergodic chain is connected with the matrix $\mathbf{R}$, with a transpose of form

$$\mathbf{R'} = \begin{pmatrix} p_{11}e^{t_{11}} & p_{12}e^{t_{12}} & \cdots & p_{1s}e^{t_{1s}} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ p_{s1}e^{t_{s1}} & \cdots & \cdots & p_{ss}e^{t_{ss}} \end{pmatrix}, \tag{10}$$

which was introduced by Montroll (1947) and Bartlett (1951), and will later be used to obtain the moment generating function of the transition frequencies n_{ij}, previously mentioned in equation (5). It is easy to see that the latent roots $\mu_r(t)$ of this matrix, where t is the matrix of elements t_{ij} $(i, j = 1, ..., s)$ are given by the equation

$$|\mathbf{R} - \mu\mathbf{I}| = 0 \tag{11}$$

and are continuous in t. For $t = 0$, this determinantal equation becomes

$$|\mathbf{p} - \mu\mathbf{I}| = 0, \tag{12}$$

with roots $\mu_1(0), ..., \mu_s(0)$, not necessarily all distinct; we may, without loss of generality, assume that these are arranged so that their moduli are in descending order of magnitude. For a stationary chain, it is known that

$$\mu_1(0) = 1, \quad |\mu_r(0)| < 1 \quad \text{for the remaining } r = 2, 3, ..., s;$$

it follows from the continuity of the roots $\mu_1(t), ..., \mu_s(t)$, that for t in the neighbourhood of $t = 0$, we have

$$|\mu_r(t)| < |\mu_1(t)| \quad \text{for all } r = 2, 3, ..., s.$$

We prove also that for an ergodic chain, for some t in the neighbourhood of $t = 0$, $\mu_1(t)$ is not identically equal to 1. For, suppose $\mu_1(t) = 1$, then for t such that $t_{11} \neq 0$, and $t_{ij} = 0$ for all other values of i, j, the equation (11) would give

$$\begin{vmatrix} p_{11}e^{t_{11}}-1 & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22}-1 & \cdots & p_{2s} \\ \hdotsfor{4} \\ p_{s1} & \cdots & \cdots & p_{ss}-1 \end{vmatrix} = 0. \tag{13}$$

On expansion, this could be written

$$(p_{11}e^{t_{11}}-1)C_{11}+p_{12}C_{12}+\ldots+p_{1s}C_{1s} = 0,$$

where the C_{1j} are cofactors of the elements in the first row and jth column. For t_{11} small, this is

$$p_{11}t_{11}C_{11}+(p_{11}-1)C_{11}+p_{12}C_{12}+\ldots+p_{1s}C_{1s} = 0.$$

Now if $t_{11} = 0$ also, so that $t = 0$, equation (13) would give

$$(p_{11}-1)C_{11}+p_{12}C_{13}+\ldots+p_{1s}C_{1s} = 0;$$

we see, therefore, that $\mu_1(t) \equiv 1$ only if $p_{11}C_{11}t_{11} = 0$, so that $p_{11} = 0$ or $C_{11} = 0$, or both are zero. C_{11} cannot be zero, for since $\mu_1(0) = 1$ is a simple root, then

$$\left\{\frac{d}{d\mu}(|\,\mathbf{p}-\mu\mathbf{I}\,|)\right\}_{\mu=1} \neq 0,$$

or, on expansion,

$$-\{C_{11}+C_{22}+\ldots+C_{ss}\} \neq 0,$$

where C_{ii} are cofactors of the elements in the leading diagonal of (13) when $t_{11} = 0$. At least one of these C_{ii} is non-zero, and we may without loss of generality assume that C_{11} is such a non-zero cofactor. If, in addition, p_{11} is non-zero, $\mu_1(t) \neq 1$ for at least the case when $t_{11} \neq 0$ and $t_{ij} = 0$ for all other i, j.

It is possible, however, that p_{11} be zero; if this is so, then for an ergodic chain, at least two of the probabilities $p_{12}, \ldots, p_{1s}$ will be non-zero.* From equation (12) we see that in this case

$$-C_{11}+p_{12}C_{12}+\ldots+p_{1s}C_{1s} = 0,$$

so that it is clear that at least one of the cofactors C_{1r} $(r = 2, \ldots, s)$ which multiplies a non-zero probability p_{1r} is non-zero. Let t now be such that $t_{1r} \neq 0$, and the remaining t_{ij} are all zero; then in exactly the same way as before, for a small value of t_{1r}, it can be shown that $\mu_1(t) \equiv 1$ only if $p_{1r}C_{1r}t_{1r} = 0$, a condition which is absurd. So that for $p_{11} = 0$, there still exists a t for which $\mu_1(t) \neq 1$. It follows therefore that in general for an ergodic chain, for some t in the neighbourhood of $t = 0$, $\mu_1(t)$ is not identically equal to 1.

(1) Consistency of the maximum-likelihood estimator

If the estimator T obtained as a solution of (7) is to be consistent, it is necessary that if θ_0 is the real value of θ, then as $n \to \infty$,

$$\Pr\{|\,T-\theta_0\,|<\delta\} \to 1, \tag{14}$$

where δ is any small positive number.

* By an ergodic chain is usually understood one for which $0 \leqslant p_{ij} \leqslant 1$, but no $p_{ii} = 1$ $(i, j = 1, \ldots, s)$. It is possible, however, that a particular $p_{ij} = 1$ for $i \neq j$: in this case the states E_i, E_j may be contracted, and we may therefore accept that for an ergodic chain, $0 \leqslant p_{ij} < 1$ for all $i, j = 1, \ldots, s$. Our statement follows immediately.

Consider the expansion

$$\frac{dL}{d\theta} = \left|\frac{dL}{d\theta}\right|_{\theta_0} + (\theta-\theta_0)\left|\frac{d^2L}{d\theta^2}\right|_{\theta_0} + \tfrac{1}{2}(\theta-\theta_0)^2\left|\frac{d^3L}{d\theta^3}\right|_{\theta_1}, \tag{15}$$

where θ_1 lies in (θ,θ_0). For simplicity, we write $\left|\frac{dL}{d\theta}\right|_{\theta_0}, \left|\frac{d^2L}{d\theta^2}\right|_{\theta_0}, \left|\frac{d^3L}{d\theta^3}\right|_{\theta_1}$ as nB_0, nB_1, nB_2 respectively, where these are obtained from (6) as

$$
\begin{aligned}
nB_0 &= \left\{\sum_{i,j}\frac{n_{ij}}{p_{ij}}\frac{dp_{ij}}{d\theta}\right\}_{\theta_0}, \\
nB_1 &= \left\{\sum_{i,j}n_{ij}\left[\frac{1}{p_{ij}}\frac{d^2p_{ij}}{d\theta^2} - \frac{1}{p_{ij}^2}\left(\frac{dp_{ij}}{d\theta}\right)^2\right]\right\}_{\theta_0}, \\
nB_2 &= \left\{\sum_{i,j}n_{ij}\left[\frac{1}{p_{ij}}\frac{d^3p_{ij}}{d\theta^3} - \frac{3}{p_{ij}^2}\frac{dp_{ij}}{d\theta}\frac{d^2p_{ij}}{d\theta^2} + \frac{2}{p_{ij}^3}\left(\frac{dp_{ij}}{d\theta}\right)^3\right]\right\}_{\theta_1}.
\end{aligned}
\tag{16}
$$

It is simple to see that as $n \to \infty$, the expectations of the n_{ij}, irrespective of whether the chain is initially or only finally stationary, will be given for all $i,j = 1, 2, ..., s$, by

$$\mathscr{E}(n_{ij}) = nP_ip_{ij}.$$

Further, since on differentiation with respect to θ, equation (1) gives

$$\sum_{j=1}^{s}\frac{dp_{ij}}{d\theta} = 0, \quad \sum_{j=1}^{s}\frac{d^2p_{ij}}{d\theta^2} = 0, \quad \sum_{j=1}^{s}\frac{d^3p_{ij}}{d\theta^3} = 0,$$

we obtain for the expectations of nB_0, nB_1, nB_2, the results

$$
\begin{aligned}
n\mathscr{E}(B_0) &= n\left\{\sum_i P_i \sum_j \frac{dp_{ij}}{d\theta}\right\}_{\theta_0} = 0, \\
n\mathscr{E}(B_1) &= -n\left\{\sum_i P_i \sum_j \frac{1}{p_{ij}}\left(\frac{dp_{ij}}{d\theta}\right)^2\right\}_{\theta_0} = -i(\theta_0)n, \\
n\mathscr{E}(B_2) &= n\left\{\sum_i P_i \sum_j \left[\frac{2}{p_{ij}^2}\left(\frac{dp_{ij}}{d\theta}\right)^3 - \frac{3}{p_{ij}}\frac{dp_{ij}}{d\theta}\frac{d^2p_{ij}}{d\theta^2}\right]\right\}_{\theta_1} = nK(\theta_1),
\end{aligned}
\tag{17}
$$

where $i(\theta)$ is obviously finite and positive. We now prove that the variates $n_{ij}n^{-1}$ converge in probability to their expectations; from this it will follow that although the $n_{ij}n^{-1}$ are not independent, B_0, B_1, B_2, which are linear functions of these variates, also jointly converge in probability to their expectations.

In order to prove the convergence in probability of the variates $n_{ij}n^{-1}$ to their expectations P_ip_{ij}, we shall evaluate their variances using a method described in Fréchet (1952, p. 73). A transition frequency n_{ij} is treated as the sum of n variates $X_{ij}^{(r)}$, so that

$$n_{ij} = \sum_{r=1}^{n} X_{ij}^{(r)},$$

where $X_{ij}^{(r)}$ takes the values 1 or 0 at the rth transition, depending on whether this transition is or is not from the state E_i to the state E_j. It is clear that

$$
\begin{aligned}
\mathscr{E}(X_{ij}^{(r)}) &= P_ip_{ij}, \\
\mathscr{V}(X_{ij}^{(r)}) &= \mathscr{E}\{(X_{ij}^{(r)})^2\} - \{\mathscr{E}(X_{ij}^{(r)})\}^2 \\
&= P_ip_{ij} - P_i^2p_{ij}^2 \quad (r = 1, 2, ..., n),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathscr{C}(X_{ij}^{(r)}X_{ij}^{(t)}) &= \mathscr{E}(X_{ij}^{(r)}X_{ij}^{(t)}) - \mathscr{E}(X_{ij}^{(r)})\mathscr{E}(X_{ij}^{(t)}) \\
&= P_ip_{ij}^2p_{ji}^{(t-r-1)} - P_i^2p_{ij}^2 \quad (r, t = 1, 2, ..., n; \ t > r),
\end{aligned}
$$

where p_{ji}^{t-r-1} is the probability that a transition from the state E_j to the state E_i occur in $t-r-1$ steps, and is the (j,i)th element of the matrix $\mathbf{p}^{t-r-1}$, the $(t-r-1)$th power of the matrix $\mathbf{p}$.

The variance σ_{ij}^2 of n_{ij} is therefore given by

$$\sigma_{ij}^2 = \mathscr{V}\left(\sum_{r=1}^{n} X_{ij}^{(r)}\right)$$

$$= \sum_{r=1}^{n} \mathscr{V}(X_{ij}^{(r)}) + 2\sum_{t>r} \mathscr{C}(X_{ij}^{(r)} X_{ij}^{(t)})$$

$$= n(P_i p_{ij} - P_i^2 p_{ij}^2) + 2P_i p_{ij}^2 \sum_{t>r} (p_{ji}^{t-r-1} - P_i),$$

where there are $\frac{1}{2}n(n-1)$ terms of the form $(p_{ji}^{t-r-1} - P_i)$. Now, since the chain is ergodic, and therefore stationary, it is known that for all $t-r-1 \geqslant n_0$, where n_0 is some finite positive integer, the terms

$$|p_{ji}^{t-r-1} - P_i|$$

converge to zero at least as fast as the terms of a convergent geometric progression with a ratio q where $0 < q < 1$ depends on n_0. If, therefore, we write the variance σ_{ij}^2 as

$$\sigma_{ij}^2 = n(P_i p_{ij} - P_i^2 p_{ij}^2) + 2P_i p_{ij}^2 \sum_{k=0}^{n-2} (n-k-1)(p_{ji}^{(k)} - P_i),$$

we see that $\sigma_{ij}^2 n^{-1}$ will clearly converge to

$$\lim_{n\to\infty} \sigma_{ij}^2 n^{-1} = P_i p_{ij} - P_i^2 p_{ij}^2 + 2P_i p_{ij}^2 s_{ji} = A,$$

where A is some value independent of n, and $s_{ji} = \sum_{k=0}^{\infty} (p_{ji}^{(k)} - P_i)$.

We now prove that for ergodic chains, the limit A must be non-zero. To do this, we use a theorem given in Fréchet (1952, pp. 86–8) which applies to the frequencies n_i of (9), giving the number of times in a realization that the system is in state E_i. The variances $\sigma_i^2 = \mathscr{V}(n_i)$ of these can be shown, by a method similar to that used for the n_{ij} above, to be such that $\lim_{n\to\infty} \sigma_i^2 n^{-1}$ is some finite value independent of n. The theorem proves that for an ergodic chain of the type we consider, $\lim_{n\to\infty} \sigma_i^2 n^{-1}$ cannot be zero.

In order to apply this result to our transition frequencies n_{ij} we redefine our system of states in the following fashion, assuming first that the p_{ij} are all non-zero. We define a system of s^2 states E_{ij} $(i, j = 1, 2, ..., s)$, in which the system will be in state E_{ij} when there is a transition from state E_i to state E_j of the original system; the new stochastic matrix for this system will be

$$\gamma = \begin{pmatrix} p_{11} \cdots p_{1s} & 0 \cdots\cdots & \cdots\cdots & \cdots\cdots 0 \\ 0 \cdots 0 & p_{21} \cdots p_{2s} & 0\cdots\cdots & \cdots\cdots 0 \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 0 \cdots\cdots & \cdots\cdots & \cdots\cdots 0 & p_{s1} \cdots p_{ss} \end{pmatrix},$$

If the original stochastic matrix is written as

$$\mathbf{p} = (\alpha_1 \alpha_2 \ldots \alpha_s) = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_s \end{pmatrix},$$

where α_i, β_i are respectively column and row vectors, then γ can be written

$$\gamma = \begin{pmatrix} \begin{pmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \beta_s \end{pmatrix} \\ \begin{Bmatrix} \\ \end{Bmatrix} \\ \cdots\cdots\cdots\cdots\cdots \\ \begin{Bmatrix} \\ \end{Bmatrix} \end{pmatrix}.$$

It is clear that n_{ij} will now indicate the number of times in a realization of the chain that the system is in the state E_{ij}; in other words, the n_{ij} in the new system are the analogues of the n_i in the original system. All that remains to be proved is that if the original system is ergodic, so also will the new system; this is intuitively obvious, and can be easily shown by powering the matrix γ as follows. If we write for the nth power of the matrix $\mathbf{p}$

$$\mathbf{p}^n = (\alpha_1^{(n)} \alpha_2^{(n)} \ldots \alpha_s^{(n)}),$$

it is seen directly that on multiplying the matrix γ by itself, we obtain

$$\gamma^2 = \begin{pmatrix} p_{11}\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_1 \end{pmatrix} \cdots p_{1s}\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_1 \end{pmatrix} & p_{21}\begin{pmatrix} \alpha_2 \\ \vdots \\ \alpha_2 \end{pmatrix} \cdots p_{2s}\begin{pmatrix} \alpha_2 \\ \vdots \\ \alpha_2 \end{pmatrix} & \cdots\cdots & p_{s1}\begin{pmatrix} \alpha_s \\ \vdots \\ \alpha_s \end{pmatrix} \cdots p_{ss}\begin{pmatrix} \alpha_s \\ \vdots \\ \alpha_s \end{pmatrix} \end{pmatrix},$$

and similarly, since for $n \to \infty$, $\displaystyle\lim_{n \to \infty} \alpha_1^{(n)} = \begin{pmatrix} P_1 \\ \vdots \\ P_1 \end{pmatrix}$,

then
$$\gamma^{n+1} = \begin{pmatrix} p_{11}\begin{pmatrix} \alpha_1^{(n)} \\ \vdots \\ \alpha_1^{(n)} \end{pmatrix} \cdots p_{ss}\begin{pmatrix} \alpha_s^{(n)} \\ \vdots \\ \alpha_s^{(n)} \end{pmatrix} \end{pmatrix} \to \begin{pmatrix} p_{11}P_1 & \cdots & p_{ss}P_s \\ \vdots & & \vdots \\ p_{11}P_1 & \cdots & p_{ss}P_s \end{pmatrix}.$$

It follows therefore by Fréchet's theorem that $\lim_{n \to \infty} \sigma_{ij}^2 \, n^{-1} \neq 0$. If a certain p_{ij} is zero, then in our new system the state E_{ij} must be eliminated, since no transition into or from this state is possible. It can be verified, however, that the results above will hold equally well in such a case. We now conclude from Tchebyshev's theorem that the variates $n_{ij} n^{-1}$ converge in probability to their expectations since

$$\Pr\{|n_{ij} n^{-1} - P_i p_{ij}| \geqslant \epsilon\} \leqslant \sigma_{ij}^2 / n^2 \epsilon^2 \leqslant A / n\epsilon^2.$$

We may express the joint convergence in probability of B_0, B_1, B_2 to their expectations in the form

$$\Pr\{|B_0| < \delta^2, \; B_1 < -\tfrac{1}{2}i(\theta_0), \; |B_2| < 2|K(\theta_1)|\} > 1 - \epsilon,$$

where δ, ϵ are any two small positive numbers. Let F denote the set of points for which this inequality holds; then for every point in F, we have that

$$B_0 + \tfrac{1}{2}B_2\delta^2 < (1 + |K(\theta_1)|)\delta^2$$

and
$$B_1\delta < -\tfrac{1}{2}i(\theta_0)\,\delta.$$

If $\theta = \theta_0 \pm \delta$ are two such points in F, then from (15) we have for these points

$$\frac{1}{n}\frac{dL}{d\theta} = B_0 \pm \delta B_1 + \tfrac{1}{2}\delta^2 B_2.$$

Provided that δ is so defined that

$$\delta < \tfrac{1}{2}i(\theta_0)(1 + |K(\theta_1)|)^{-1},$$

then for $\theta = \theta_0 + \delta$,

$$\frac{1}{n}\frac{dL}{d\theta} < (1 + |K(\theta_1)|)\{\delta - \tfrac{1}{2}i(\theta_0)(1 + |K(\theta_1)|)^{-1}\}\delta < 0.$$

whereas for $\theta = \theta_0 - \delta$,

$$\frac{1}{n}\frac{dL}{d\theta} > (1 + |K(\theta_1)|)\{\delta + \tfrac{1}{2}i(\theta_0)(1 + |K(\theta_1)|)^{-1}\}\delta > 0.$$

We see from (6) that $dL/d\theta$ is a continuous function of θ in R, and it follows therefore that the likelihood equation (7) has a root in $\theta_0 \pm \delta$ with probability tending to 1. This establishes the consistency of the estimator T of θ.

(2) *Asymptotic normality of maximum-likelihood estimator*

Let the consistent solution of the likelihood equation be T; then from equation (15), we have that

$$\frac{1}{n}\left|\frac{dL}{d\theta}\right|_T = \frac{1}{n}\left|\frac{dL}{d\theta}\right|_{\theta_0} + (T - \theta_0)\frac{1}{n}\left|\frac{d^2L}{d\theta^2}\right|_{\theta_0} + \tfrac{1}{2}(T - \theta_0)^2\frac{1}{n}\left|\frac{d^3L}{d\theta^3}\right|_{\theta_1} = 0,$$

or

$$B_0 + (T - \theta_0)B_1 + \tfrac{1}{2}(T - \theta_0)^2 B_2 = 0,$$

so that

$$(T - \theta_0) = -B_0\{B_1 + \tfrac{1}{2}(T - \theta_0)B_2\}^{-1}.$$

We have seen that the expectation of B_0 is zero; its variance can be evaluated, again using the method described in Fréchet (1952, p. 73):

$$\mathscr{V}(B_0) = \mathscr{V}\left\{\sum_{i,j} n_{ij} n^{-1}\left(\frac{1}{p_{ij}}\frac{dp_{ij}}{d\theta}\right)_T\right\}$$

$$= \frac{1}{n^2}\mathscr{E}\left\{\sum_{i,j}\left(\frac{n_{ij}}{p_{ij}}\frac{dp_{ij}}{d\theta}\right)_T\right\}^2. \tag{18}$$

We treat each n_{ij} as the sum of the n variates $X_{ij}^{(r)}$ which take the values 1 or 0 at each of the n transitions $r = 1, \ldots, n$, depending on whether these transitions are or are not from the state E_i to the state E_j.

We then have

$$\mathscr{E}\left\{\sum_{i,j}\frac{n_{ij}}{p_{ij}}\frac{dp_{ij}}{d\theta}\right\}^2 = \mathscr{E}\left\{\sum_r\left(\sum_{i,j}\frac{X_{ij}^{(r)}}{p_{ij}}\frac{dp_{ij}}{d\theta}\right)\right\}^2$$

$$= \sum_r\mathscr{E}\left\{\sum_{i,j}\frac{X_{ij}^{(r)}}{p_{ij}}\frac{dp_{ij}}{d\theta}\right\}^2 + 2\sum_{r<t}\mathscr{E}\left\{\sum_{i,j}\frac{X_{ij}^{(r)}}{p_{ij}}\frac{dp_{ij}}{d\theta}\sum_{l,m}\frac{X_{lm}^{(t)}}{p_{lm}}\frac{dp_{lm}}{d\theta}\right\}$$

$$= \sum_r\mathscr{E}\left\{\sum_{i,j}\frac{(X_{ij}^{(r)})^2}{p_{ij}^2}\left(\frac{dp_{ij}}{d\theta}\right)^2 + \sum'\sum_{i,j\ l,m}\frac{X_{ij}^{(r)}X_{lm}^{(r)}}{p_{ij}\ p_{lm}}\frac{dp_{ij}}{d\theta}\frac{dp_{lm}}{d\theta}\right\}$$

$$+ 2\sum_{r<t}\left\{\sum_{i,j}\sum_{l,m}\mathscr{E}\left(\frac{X_{ij}^{(r)}X_{lm}^{(t)}}{p_{ij}\ p_{lm}}\frac{dp_{ij}}{d\theta}\frac{dp_{lm}}{d\theta}\right)\right\},$$

where $\sum\limits_{r}\sum\limits_{i,j}\sum\limits_{l,m}$ indicate summation over all possible values $r = 1, \ldots, n; i, j$ and $l, m = 1, \ldots, s$, but where $\sum\limits_{i,j}'\sum\limits_{l,m}'$ indicate summation over those values of i, j and $l, m = 1, \ldots, s$, for which at least one of the conditions $i \neq l, j \neq m$ holds. Since for these values of i, j and l, m, we have that $X_{ij}^{(r)} X_{lm}^{(r)} = 0$, and further $\sum\limits_{m} \dfrac{dp_{lm}}{d\theta} = 0$ for all l, then this is equal to

$$\sum_{i,j} \mathscr{E}\left\{\sum_{r} \frac{X_{ij}^{(r)}}{p_{ij}^2}\left(\frac{dp_{ij}}{d\theta}\right)^2\right\} + 2 \sum_{r<t}\left\{\sum_{i,j\,l,m}(P_i p_{ij} p_{jl}^{(t-r-1)} p_{lm})\left(\frac{1}{p_{ij}p_{lm}}\cdot\frac{dp_{ij}}{d\theta}\frac{dp_{lm}}{d\theta}\right)\right\} = n\sum_{i,j}\frac{P_i}{p_{ij}}\left(\frac{dp_{ij}}{d\theta}\right)^2.$$

It follows that
$$\mathscr{V}(B_0) = -n^{-1}\mathscr{E}(B_1) = n^{-1}i(\theta_0),$$

so that

$$(T - \theta_0)\,\sqrt{\{i(\theta_0)\,n\}} = \frac{\dfrac{B_0}{\sqrt{\{n^{-1}i(\theta_0)\}}}}{\left|\dfrac{B_1}{-i(\theta_0)} + \dfrac{(T-\theta_0)\,B_2}{-2i(\theta_0)}\right|}$$

has zero expectation and unit variance, for the denominator converges to 1 in probability, and the numerator has zero expectation and unit variance. If we can further show that B_0 is asymptotically normally distributed, then T will be asymptotically normally distributed with expectation θ_0 and variance $\{i(\theta_0)\,n\}^{-1}$.

Now from (16) we see that B_0 is a linear function of the n_{ij}, so that it is sufficient to prove that as $n \to \infty$, these are jointly normally distributed. This has been proved, for the ergodic chains considered, by Bartlett (1951), who in his work also implied some of the properties of maximum-likelihood estimators developed in detail in this paper. The proof is briefly outlined below; one difference, however, is that it is no longer necessary to postulate $\ln \mu_1(t) \neq 0$, since it was shown in § III that for the chains considered, this must be so for some t.

Following Montroll (1947) and Bartlett (1951), we define the moment generating function of the n_{ij} as

$$M(t) = \mathscr{E}(\exp \sum_{i,j} t_{ij} n_{ij})$$
$$= \mathscr{E}(\exp \sum_{i,j} t_{ij} \sum_{r} X_{ij}^{(r)}).$$

If $\mathbf{R}$ is the square matrix whose transpose $\mathbf{R}'$ is defined by (10), this is

$$M(t) = \mathbf{1}'\mathbf{R}^n\mathbf{P},$$

where $\mathbf{1}'$ denotes the row vector $(1, \ldots, 1)$, and $\mathbf{P}$ is the column vector of the stationary probabilities P_i. If $\mathbf{R}$ has the s latent roots $\mu_1(t), \ldots, \mu_s(t)$, we have seen that for small t, $\mu_1(t)$ is the single dominant root such that $\mu_1(0) = 1$. Following Frazer, Duncan & Collar (1947, §4·15), we have that Sylvester's theorem gives for $n \to \infty$,

$$\mathbf{R}^n \sim \{\mu_1(t)\}^n\, \mathbf{Z}_0(\mu_1),$$

where the matrix $\mathbf{Z}_0(\mu_1)$ is finite. It follows that

$$M(t) \sim \{\mu_1(t)\}^n\, \mathbf{1}'\mathbf{Z}_0(\mu_1)\,\mathbf{P},$$

so that
$$\ln M(t) \sim n \ln \mu_1(t),$$

a non-zero value, since for an ergodic chain, for some small t it has been shown that $\mu_1(t) \neq 1$.

$M(t)$ is thus asymptotically equivalent to the moment generating function of a sum of n identical independent vector variates. Since for an ergodic chain, the n_{ij} have variances of order n so that $\sigma_{ij}^2 n^{-1}$ are finite and non-zero, they will tend to simultaneous normality as $n \to \infty$ by the Central Limit Theorem for $n = \sum_{i,j} n_{ij}$ variates. It follows that B_0 will be asymptotically normally distributed.

Finally, a general convergence theorem of Cramér (1946) enables us to conclude that $(T - \theta_0) \sqrt{\{i(\theta_0) n\}}$ is asymptotically normally distributed with zero expectation and unit variance; this is that if a variate ξ_n with distribution function $f_n(x)$ tending to $f(x)$ and a variate η_n converging in probability to a positive constant c as $n \to \infty$ exist, then the distribution function of $\xi_n \eta_n^{-1}$ tends to $f(cx)$.

Since
$$\{i(\theta_0)n\}^{-1} = \left\{ \mathscr{E} \left(\left| \frac{d^2 L}{d\theta^2} \right|_{\theta_0} \right) \right\}^{-1}$$

is the minimum possible variance, we see clearly that the consistent solution of the maximum-likelihood equation is fully efficient.

IV. Sufficiency conditions

A further general theorem of maximum-likelihood estimation, proved under certain conditions by Cramér (1946, §33·2) for continuous distributions, applies equally with minor modifications to the case of the Markov chain. It is that:

If a sufficient estimator T of the parameter θ exists, any solution of the likelihood equation will be a function of T. In obtaining the maximum-likelihood estimator T of θ, the parameter defining the transition probabilities $p_{ij}(\theta)$ of a Markov chain, it may therefore be of interest to determine the form of the transition probabilities which admit a sufficient statistic for θ.

Koopman (1936) and Pitman (1936) have given the general form of continuous distributions admitting a sufficient statistic, but although it has been obvious that a similar form exists for discrete distributions, no account of the proof for this appears to be available. We briefly outline the proof, use it to obtain the form of the probabilities $p_1(\theta), \ldots, p_k(\theta)$, in the particular case of the multinomial distribution, and finally generalize our results in the case of the simple ergodic Markov chain.

(1) *The form of the likelihood function admitting a sufficient estimator T of θ,*
for discrete parent distributions with constant variate intervals

We consider those discrete distributions $p(x, \theta)$ for which the variate x is defined at equal intervals. We can then, without loss of generality, assume the interval to be unity so that x takes consecutive integral values in any specific range, finite or infinite; this includes the standard discrete distributions for which x is a frequency, as well as others for which x may take positive or negative integral values.

Let $x_1, \ldots, x_n$ be n independent observations of a variate x with the discrete distribution $p(x, \theta)$, where x takes a set of consecutive integral values in a given range. The probability of this sample is
$$\phi(x_1, \ldots, x_n; \theta) = p(x_1, \theta) \ldots p(x_n, \theta).$$

If this is to admit a sufficient estimator T of θ, then the factorizability condition
$$\phi(x_1, \ldots, x_n; \theta) = f(x_1, \ldots, x_n) F(T, \theta)$$

must hold, or for the likelihood function $L = \ln \phi$,

$$L = \ln f(x_1, ..., x_n) + \ln F(T, \theta). \tag{19}$$

Assuming that the functions $p(x_i, \theta)$, $F(T, \theta)$ are differentiable with respect to θ, we obtain the equation

$$\sum_i \frac{\partial}{\partial \theta} \{\ln p(x_i, \theta)\} = \frac{\partial}{\partial \theta} \{\ln F(T, \theta)\} = G(T, \theta). \tag{20}$$

Since this holds for all values of θ within the range allowed, a particular value θ_0 lying in this range will give, when substituted in this equation,

$$u = \sum_i u(x_i) = g(T), \tag{21}$$

which connects T with the statistic

$$u = \sum_i u(x_i) = \sum_i \left(\frac{\partial}{\partial \theta} \{\ln p(x_i, \theta)\} \right)_{\theta = \theta_0}.$$

Now since sufficiency is a property which holds for all values of the sample $x_1, ..., x_n$, we may allow x_i, a particular discrete sample value, to increase by 1, while the remaining $x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_n$, remain unchanged. We can then difference equation (21) with respect to x_i to obtain

$$\Delta_{x_i} u = \Delta_{x_i} u(x_i) = \Delta_{x_i} g(T).$$

Similarly for equation (20)

$$\Delta_{x_i} \frac{\partial}{\partial \theta} \ln p(x_i, \theta) = \Delta_{x_i} G(T, \theta),$$

so that for all values of i, we have the equation

$$\frac{\Delta_{x_i} \frac{\partial}{\partial \theta} \ln p(x_i, \theta)}{\Delta_{x_i} u(x_i)} = \frac{\Delta_{x_i} G(T, \theta)}{\Delta_{x_i} g(T)}. \tag{22}$$

Now if $G(T, \theta)$ and $g(T)$ are assumed differentiable with respect to T, we have that

$$\begin{aligned}
\Delta_{x_i} G(T, \theta) &= G(T(x_1, ..., x_{i-1}, x_i + 1, x_{i+1}, ..., x_n), \theta) - G(T(x_1, ..., x_i, ..., x_n), \theta) \\
&= G(T_1, \theta) - G(T_0, \theta) \\
&= (T_1 - T_0) \left\{ \frac{\partial G}{\partial T} \right\}_{T_2} \\
&= (\Delta_{x_i} T) \left\{ \frac{\partial G}{\partial T} \right\}_{T_2},
\end{aligned}$$

where T_2 is some function $T_2(x_1, ..., x_i, ..., x_n; \theta)$ of the sample values and of θ such that $T_0 \leqslant T_2 \leqslant T_1$. Further, since the estimator T is a symmetrical function of the x_i, the function T_2 will be the same in the case of every differencing with respect to each x_i, $i = 1, ..., n$. Similarly,

$$\Delta_{x_i} g(T) = (\Delta_{x_i} T) \left\{ \frac{\partial g}{\partial T} \right\}_{T_2(x_1, ..., x_i, ..., x_n; \theta_0)},$$

where

$$T_0 \leqslant T_2(x_1, ..., x_i, ..., x_n; \theta_0) \leqslant T_1.$$

It follows that equation (22) can be written for all values of i as

$$\frac{\Delta_{x_i} \frac{\partial}{\partial \theta} \ln p(x_i, \theta)}{\Delta_{x_i} u(x_i)} = \frac{\left\{ \frac{\partial G}{\partial T} \right\}_{T_2(x_1, ..., x_n; \theta)}}{\left\{ \frac{dg}{dT} \right\}_{T_2(x_1, ..., x_n; \theta_0)}},$$

which is equal to a fixed value, not depending on i. Hence

$$\frac{\Delta_{x_1}\frac{\partial}{\partial \theta}\ln p(x_1,\theta)}{\Delta_{x_1}u(x_1)} = \frac{\Delta_{x_2}\frac{\partial}{\partial \theta}\ln p(x_2,\theta)}{\Delta_{x_2}u(x_2)} = \dots = \frac{\Delta_{x_n}\frac{\partial}{\partial \theta}\ln p(x_n,\theta)}{\Delta_{x_n}u(x_n)}$$

and this can only be so if these are all equal to a function of θ only, $K_1(\theta)$. It follows from (22) that

$$\Delta_{x_i}G(T,\theta) = K_1(\theta)\,\Delta_{x_i}g(T),$$

so that

$$G(T,\theta) = K_1(\theta)\,g(T) + K_2(\theta).$$

On integrating this with respect to θ, we obtain

$$\ln F(T,\theta) = \Lambda_1(\theta)\,g(T) + \Lambda_2(\theta) + h(T),$$

so that the likelihood (19) is of the form

$$L = \ln f(x_1,\dots,x_n) + \Lambda_1(\theta)\,g(T) + \Lambda_2(\theta), \tag{23}$$

and the probability of the sample is of the form

$$\phi(x_1,\dots,x_n;\theta) = f(x_1,\dots,x_n)\exp[\Lambda_1(\theta)\,g(T) + \Lambda_2(\theta)].$$

(2) Form of the probabilities $p_i(\theta)$ for a multinomial distribution admitting a sufficient estimator T of θ

Consider the k mutually exclusive events $E_1, E_2, \dots, E_k$, with non-zero probabilities $p_1(\theta), \dots, p_k(\theta)$ such that $\sum_{i=1}^{k} p_i(\theta) = 1$. In a total of n independent trials in which there are x_1 occurrences of the events E_1, x_2 of the event $E_2, \dots$, and x_k of the event E_k, the probability of the sample of frequencies $x_1, x_2, \dots, x_k$, is

$$\phi(x_1,\dots,x_k;\theta) = p_1^{x_1}(\theta)\dots p_k^{x_k}(\theta).$$

For this to admit a sufficient estimator, the likelihood $L = \ln\phi$ must be of the form (23), so that

$$L = \sum_{i=1}^{k} x_i \ln p_i(\theta) = \ln f(x_1,\dots,x_k) + \Lambda_1(\theta)\,g(T) + \Lambda_2(\theta).$$

We note in this case that $\Lambda_2(\theta)$ involves the number of trials $n = \sum_{i=1}^{k} x_i$ so that it can be written

$$\Lambda_2(\theta) = \sum_{i=1}^{k} x_i \lambda_2(\theta),$$

where $\lambda_2(\theta)$ is a function of θ only. Also from the form of the likelihood, we see that the first term can only be of the form

$$\ln f(x_1,\dots,x_k) = \sum_{i=1}^{k} A_i x_i.$$

We can therefore write the likelihood function admitting a sufficient estimator T of θ as

$$L = \sum_i x_i \ln p_i(\theta) = \sum_i A_i x_i + \Lambda_1(\theta)\,g(T) + \lambda_2(\theta)\sum_i x_i, \tag{24}$$

which holds for all values of n, and of the frequency values $x_1, \dots, x_n$. If, in a total of $n+1$ independent trials, the frequency of each event $E_1, E_2, \dots, E_{i-1}, E_{i+1}, \dots, E_k$ apart from the

event E_i remains unchanged, while the event E_j occurs $x_i + 1$ times, then we may difference equation (24) with respect to x_i, to obtain

$$\ln p_i(\theta) = A_i + \Lambda_1(\theta)\,\Delta_{z_i} g(T) + \lambda_2(\theta),$$

or

$$\{\Lambda_1(\theta)\}^{-1}\{\ln p_i(\theta) - A_i - \lambda_2(\theta)\} = \Delta_{z_i} g(T).$$

That is, a function of θ only equals a function of the sample values x_i, which is possible only if both equal a constant K_i.

It follows that

$$p_i(\theta) = \exp[K_i \Lambda_1(\theta) + \lambda_2(\theta) + A_i]$$
$$= \alpha_i \exp[K_i \Lambda_1(\theta) + \lambda_2(\theta)] \quad (i = 1, ..., k), \tag{25}$$

where $A_i = \ln \alpha_i$, is the form of the probability allowing a sufficient estimator T of θ.

A condition to which the probabilities $p_i(\theta)$ are subject is

$$\sum_{i=1}^{k} p_i(\theta) = 1,$$

so that in our case, from (25)

$$\sum_i \alpha_i \exp[K_i \Lambda_1(\theta) + \lambda_2(\theta)] = 1,$$

or

$$\sum_i \alpha_i \exp[K_i \Lambda_1(\theta)] = \exp[-\lambda_2(\theta)].$$

An alternative way of writing the probabilities $p_i(\theta)$ is therefore

$$p_i(\theta) = \alpha_i \exp[K_i \Lambda_1(\theta)]\{\sum_i \alpha_i \exp[K_i \Lambda_1(\theta)]\}^{-1}$$
$$= \alpha_i \lambda_1^{K_i}(\theta)\{\sum_i \alpha_i \lambda_1^{K_i}(\theta)\}^{-1}, \tag{26}$$

where

$$\lambda_1(\theta) = \exp[\Lambda_1(\theta)].$$

We can now tell immediately in any practical case whether it is possible to obtain a sufficient estimator or not by an examination of the form of the multinomial probabilities. Three simple illustrations from trinomial distributions are given as examples.

Example IV 2·1. The trinomial distribution with probabilities

$$p_1 = \theta\{\theta + \theta^2 + 2\theta^3\}^{-1}, \quad p_2 = \theta^2\{\theta + \theta^2 + 2\theta^3\}^{-1}, \quad p_3 = 2\theta^3\{\theta + \theta^2 + 2\theta^3\}^{-1},$$

of the form (26) with α_i equal to 1, 1, 2 respectively and K_i equal to 1, 2, 3, where $\lambda_1(\theta) = \theta$, will give an estimator of θ which is sufficient.

Example IV. 2·2. Probabilities of the type

$$p_1 = \theta, \quad p_2 = 2\theta, \quad p_3 = 1 - 3\theta,$$

which can be written in the form (25)

$$p_i = \alpha_i \exp[K_i \ln \theta(1 - 3\theta)^{-1} + \ln(1 - 3\theta)],$$

where α_i equal 1, 2, 1, and K_i equal 1, 1, 0, respectively, will also give an estimator of θ which is sufficient.

Example IV. 2·3. However, probabilities of the form

$$p_1 = \tfrac{1}{4}(2 + \theta), \quad p_2 = \tfrac{1}{4}(1 - \theta), \quad p_3 = \tfrac{1}{4}\theta,$$

which occur in genetics, cannot be written in the forms (25) or (26), and can be verified to admit no sufficient estimator of θ.

(3) *Form of the transition probabilities $p_{ij}(\theta)$ for a simple ergodic Markov chain admitting a sufficient estimator T of θ*

Consider the realization of the simple ergodic Markov chain with s possible states $E_1, \ldots, E_s$ which results in the observed sequence S of the $n+1$ states $E_i, E_j, \ldots E_k, E_l$ in this order. Assuming the initial state E_i to be fixed,[*] the probability distribution of the sequence S is given by (2), and the likelihood can be written as equation (3). An ergodic chain (see footnote, § III) is such that for any row i of the transition matrix, and for all values of j,

$$0 \leqslant p_{ij}(\theta) < 1,$$

so that, in particular, for $i = j$ no state E_i is an absorption state; some, though not all, $p_{ij}(\theta)$ in a row may equal zero, subject to the usual conditions (1) that

$$\sum_{j=1}^{s} p_{ij}(\theta) = 1 \quad (i = 1, 2, \ldots, s).$$

Now if the chain admits a sufficient estimator T of θ, we see that it may be written, much in the same way as for the multinomial case,

$$L = \sum_{i,j} n_{ij} \ln p_{ij}(\theta) = \sum_{i,j} A_{ij} n_{ij} + \Lambda_1(\theta) g(T) + \lambda_2(\theta) \sum_{i,j} n_{ij}, \tag{27}$$

where the A_{ij} are constants, $\Lambda_2(\theta)$ is equal to the function $\lambda_2(\theta) \sum_{i,j} n_{ij}$ in this case, and $\sum_{i,j} n_{ij} = n$.

Before proceeding, as in the similar case of the multinomial, to increase the number of trials in the realization of the chain by one, so as to allow this equation to be differenced with respect to the n_{ij} in order to obtain the non-zero values of the $p_{ij}(\theta)$, we must consider some difficulties which this method presents. Since to any zero transition probability $p_{ij}(\theta)$ there will necessarily correspond a zero value of the transition frequency n_{ij}, neither will appear in the likelihood function (27). We are concerned with the n_{ij} corresponding to non-zero values of $p_{ij}(\theta)$, and differencing occurs only with respect to these; however, owing to the set of relations (8) for the sums of transition frequencies, it is not always possible in increasing the number of trials in a realization of the chain by 1, from $n+1$ to $n+2$, to increase each of the n_{ij} associated with a non-zero transition probability $p_{ij}(\theta)$ singly by 1, while leaving the values of the remaining n_{ij} unchanged.

As a simple example to illustrate this point, we consider the case of a chain with two states E_1 and E_2, for which a possible realization of three trials, starting with the fixed state E_1, results in the sequence E_1, E_1, E_2. Assuming all the transition probabilities to

---

[*] We are specifically considering the case of sufficiency for a distribution conditional on the given initial state E_i. In the general case where any state E_i may occur initially, the likelihood is

$$L = \sum_{i=1}^{s} x_i \ln a_i + \sum_{i,j} n_{ij} \ln p_{ij},$$

where the variates x_i are 0 or 1 according to whether the state E_i is or is not the initial state. In addition to the usual conditions for the n_{ij}, the x_i are subject to the condition that

$$\sum_{i=1}^{s} x_i = 1.$$

This more difficult case will not be considered.

have non-zero values, we see immediately that the transition frequency matrix with elements n_{ij} is

$$\mathbf{n} = \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix};$$

a realization of four trials starting with E_1 will give, among others, the following transition frequency matrices:

$$\begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

for the sequences E_1, E_1, E_1, E_2; E_1, E_1, E_2, E_1; and E_1, E_1, E_2, E_2. Here we see that n_{11}, n_{21}, n_{22} have singly been increased by 1, while the remaining n_{ij} remain constant; it is, however, in no way possible, starting with the given matrix $\mathbf{n}$, to increase n_{12} by 1, while leaving n_{11}, n_{21}, n_{22} unchanged.

In order to resolve this difficulty, we use the fact that for an ergodic chain, providing the number of trials $n + 1$ is sufficiently large in a realization starting with the given state E_1, it is possible to end with any one of the states $E_1, ..., E_s$. In our particular case, assuming a sufficiently large number of trials, the realization of the chain leads to the sequence S which starts with the fixed state E_i and ends with the state E_j; a further trial may result in the sequence S' with an additional state E_r, for all r for which $p_{lr}(\theta) \neq 0$. The probability of this sequence is

$$\phi(S') = p_{ij}(\theta) \dots p_{kl}(\theta)\, p_{lr}(\theta),$$

and the likelihood function $L' = \ln \phi(S')$ is identical with $L = \ln \phi(S)$ except that the transition frequency n_{lr} in it is increased to $n_{lr} + 1$. This allows us to difference the equation (27) with respect to n_{lr} for all r for which $p_{lr} \neq 0$, in order to obtain

$$\ln p_{lr}(\theta) = A_{lr} + \Lambda_1(\theta)\, \Delta_{n_{lr}} g(T) + \lambda_2(\theta),$$

where $\Delta_{n_{lr}} g(T)$ is some function of the transition frequencies n_{ij}. Now since

$$\{\Lambda_1(\theta)\}^{-1}\{\ln p_{lr} - \lambda_2(\theta) - A_{lr}\} = \Delta_{n_{lr}} g(T),$$

that is, a function of θ equals a function of the n_{ij}, then both must equal a constant K_{lr}. This allows us to write for all r for which $p_{lr} \neq 0$,

$$p_{lr}(\theta) = \alpha_{lr} \exp\left[K_{lr} \Lambda_1(\theta) + \lambda_2(\theta) \right],$$

where $A_{lr} = \ln \alpha_{lr}$. If we further admit zero values of α_{lr}, we may accept this equation as the general form for both zero and non-zero values of p_{lr} $(r = 1, 2, ..., s)$. Since, for a sufficiently large number of trials in a realization of the chain starting with the given state E_i, the final state E_l may be such that l can take any of the values $1, 2, ..., s$, it follows that we may write as the general form of the transition probabilities $p_{ij}(\theta)$ admitting a sufficient estimator T of θ, the equation

$$p_{ij}(\theta) = \alpha_{ij} \exp\left[K_{ij} \Lambda_1(\theta) + \lambda_2(\theta) \right]. \tag{28}$$

This general form of the $p_{ij}(\theta)$ has been obtained for sufficiently large values of $n = \sum_{i,j} n_{ij}$; this means that in this case the necessary condition that a sufficient estimator of θ exists is that the $p_{ij}(\theta)$ be of form (28). It is easily verified that the sufficient condition also follows; that if the $p_{ij}(\theta)$ are of the form (28), the estimator of θ is sufficient. This, however, holds generally, irrespective of the size of n. It is of interest to note that if we restrict ourselves to the case where all the $p_{ij}(\theta)$ are non-zero, all values of $n \geq 1$ are sufficiently large for the

necessary condition to hold, since any final state k may be reached as a connection of one or more trials starting with the given state E. This means that for the ergodic chain with non-zero $p_{ij}(\theta)$ a sufficient estimator of θ exists for any number of trials and θ and any j if the $p_{ij}(\theta)$ are of the form (28).

The transition probabilities (28) are subject to the condition (1) that $\sum_{j=1}^{s} p_{ij} = 1$, so that

$$\sum_{j=1}^{s} \alpha_{ij} \exp[K_{ij}\Lambda_1(\theta)] = \exp[-\lambda_s(\theta)].$$

or $$\sum_{j=1}^{s} \alpha_{ij} \exp[K_{ij}\Lambda_1(\theta)] = \sum_{j=1}^{s} \alpha_{rj} \exp[K_{rj}\Lambda_1(\theta)] \quad (i, r = 1, 2, \ldots, s) \qquad (29)$$

If we write $\lambda_1(\theta) = \exp[\Lambda_1(\theta)]$ we see that the $p_{ij}(\theta)$ can also be put in the form

$$p_{ij}(\theta) = \alpha_{ij}\lambda_1^{K_{ij}}(\theta)\{\sum_{j} \alpha_{ij}\lambda_1^{K_{ij}}(\theta)\}^{-1}, \qquad (30)$$

where $\{\sum_{j} \alpha_{ij}\lambda_1^{K_{ij}}(\theta)\}^{-1}$ has the same value for all rows $i = 1, 2, \ldots, s$.

Two possible cases now exist for those exponents among $K_{i1}, K_{i2}, \ldots, K_{is}$, associated with the non-zero probabilities among $p_{i1}, p_{i2}, \ldots, p_{is}$, in the ith row of the transition probability matrix $\mathbf{p}$: the first in which all such exponents are distinct, and the second in which only some such exponents are distinct. For simplicity, we assume that in the ith row, all transition probabilities $p_{ij}(\theta)$ $(j = 1, 2, \ldots, s)$ are non-zero, the slightly more general case where some $p_{ij}(\theta)$ may be zero follows without difficulty.

Consider the first case, where for the particular row i there are s exponents K_{ij} all distinct, associated with the s non-zero transition probabilities $p_{ij}(\theta)$ $(j = 1, 2, \ldots, s)$, then, from equation (29), it follows that for any other row r,

$$\sum_{j} \alpha_{ij}\lambda_1^{K_{ij}} = \sum_{j} \alpha_{rj}\lambda_1^{K_{rj}}. \qquad (31)$$

This is possible only if the exponents K_{ij} and K_{rj} $(j = 1, 2, \ldots, s)$ have the same s distinct values, though these may be arranged in different orders. This means that there is only a single set of s distinct values

$$\alpha_{i1}\lambda_1^{K_{i1}}\{\sum_{j} \alpha_{ij}\lambda_1^{K_{ij}}\}^{-1}, \quad \ldots, \quad \alpha_{is}\lambda_1^{K_{is}}\{\sum_{j} \alpha_{ij}\lambda_1^{K_{ij}}\}^{-1},$$

for the transition probabilities of the matrix $\mathbf{p}$, and these must appear, possibly in different orders, in every row of the matrix.

Two simple examples for Markov chains with two and three states respectively will illustrate the previous points:

Example IV. 3·1. The transition probability matrix of the form

$$\mathbf{p} = \begin{pmatrix} \theta & 1-\theta \\ 1-\theta & \theta \end{pmatrix}$$

will always provide a sufficient estimator of θ, for we may write $p_{ij}(\theta)$ $(j = 1, 2)$ in the form (28) as

$$p_{1j}(\theta) = \alpha_{1j}\exp[K_{1j}\ln\theta(1-\theta)^{-1} + \ln(1-\theta)],$$

where the values of α_{11}, α_{12} and K_{11}, K_{12} are 1, 1 and 1, 0 respectively. The second row consists of the same transition probabilities in a different order.

Example IV. 3·2.

$$p = \begin{pmatrix} 2\theta(2\theta + \theta^2 + \theta^3)^{-1} & \theta^2(2\theta + \theta^2 + \theta^3)^{-1} & \theta^3(2\theta + \theta^2 + \theta^3)^{-1} \\ \theta^3(2\theta + \theta^2 + \theta^3)^{-1} & 2\theta(2\theta + \theta^2 + \theta^3)^{-1} & \theta^2(2\theta + \theta^2 + \theta^3)^{-1} \\ \theta^2(2\theta + \theta^2 + \theta^3)^{-1} & \theta^3(2\theta + \theta^2 + \theta^3)^{-1} & 2\theta(2\theta + \theta^2 + \theta^3)^{-1} \end{pmatrix}$$

also provides a sufficient estimator of θ, for in (30) we have $\lambda_1(\theta) = \theta$, and the values of α_{1j} and K_{1j} are 2, 1, 1 and 1, 2, 3 respectively. The second and third rows consist of the same values for the transition probabilities, but in different orders.

Consider now the second case for the exponents. Here, for the particular row i, the s exponents K_{ij} are not all distinct; let the first k exponents $(k < s)$ be identical

$$K_{i1} = K_{i2} = \dots = K_{ik} = K$$

and the remaining $K_{i,k+1}, \dots, K_{is}$ be distinct, so that there are altogether $s - k + 1$ distinct values for the exponents, and the transition probabilities for row i are therefore

$$\begin{aligned} p_{ij}(\theta) &= \alpha_{ij}\lambda_1^K \left\{ \sum_{j=1}^{s} \alpha_{ij}\lambda_1^{K_{ij}} \right\}^{-1} \quad (j = 1, 2, \dots, k), \\ p_{ij}(\theta) &= \alpha_{ij}\lambda_1^{K_{ij}} \left\{ \sum_{j=1}^{s} \alpha_{ij}\lambda_1^{K_{ij}} \right\}^{-1} \quad (j = k+1, \dots, s). \end{aligned} \tag{32}$$

(The slightly more general case of several groups of exponents with identical values, and the remaining exponents distinct

$$\begin{aligned} K_{i1} &= K_{i2} = \dots = K_{ik} = K^{(1)}, \\ K_{i,k+1} &= K_{i,k+2} = \dots = K_{im} = K^{(2)}, \\ &\quad\dots\dots\dots\dots\dots\dots\dots\dots \\ K_{iu} &\neq K_{i,u+1} \neq \dots \neq K_{is}, \end{aligned}$$

presents no difficulty different from those met in the simpler case mentioned, and will not be considered.) Now since for any row r equation (31) holds, there are $s - k + 1$ distinct values of K_{rj} which are identical with the distinct values of the K_{ij}, so that in every row of the stochastic matrix p we have, apart from coefficients, transition probabilities of only the following $s - k + 1$ distinct forms

$$\lambda_1^K \{ \textstyle\sum_j \alpha_{ij}\lambda_1^{K_{ij}} \}^{-1}, \quad \lambda_1^{K_{i,k+1}} \{ \textstyle\sum_j \alpha_{ij}\lambda_1^{K_{ij}} \}^{-1}, \quad \dots, \quad \lambda_1^{K_{is}} \{ \textstyle\sum_j \alpha_{ij}\lambda_1^{K_{ij}} \}^{-1}, \tag{33}$$

appearing in various arrangements. The s coefficients associated with these $s - k + 1$ distinct forms for the transition probabilities may also differ from row to row, but are subject to the condition which follows from (31) that for all rows $r = 1, 2, \dots, s$, the sums of the coefficients for each distinct form in (33) must equal the fixed values

$$\sum_{j=1}^{k} \alpha_{ij}, \quad \alpha_{i,k+1}, \quad \dots, \quad \alpha_{is}, \tag{34}$$

respectively.

Two simple examples of this case for Markov chains with three states follow.

Example IV. 3·3.

$$p = \begin{pmatrix} 2\theta(3\theta + \theta^3)^{-1} & \theta(3\theta + \theta^3)^{-1} & \theta^3(3\theta + \theta^3)^{-1} \\ \frac{1}{3}\theta^3(3\theta + \theta^3)^{-1} & \frac{2}{3}\theta^3(3\theta + \theta^3)^{-1} & 3\theta(3\theta + \theta^3)^{-1} \\ \frac{11}{4}\theta(3\theta + \theta^3)^{-1} & \theta^3(3\theta + \theta^3)^{-1} & \frac{1}{4}\theta(3\theta + \theta^3)^{-1} \end{pmatrix},$$

where in equation (32), $\lambda_1(\theta) = 0$, and the two distinct forms for the transition probabilities appearing in each row are

$$\theta(3\theta + \theta^2)^{-1}, \quad \theta^2(3\theta + \theta^2)^{-1}$$

The values of the coefficients for the form $\theta(3\theta + \theta^2)^{-1}$ vary from row to row, subject to the condition (34) that their sum is always equal to 3, while those for the form $\theta^2(3\theta + \theta^2)^{-1}$ have a sum always equal to 1.

Example IV. 3·4.

$$\mathbf{p} = \begin{pmatrix} \tfrac{1}{2}\theta & 1-\theta & \tfrac{1}{2}\theta \\ \tfrac{1}{4}\theta & \tfrac{3}{4}\theta & 1-\theta \\ \tfrac{1}{2}(1-\theta) & \tfrac{1}{2}(1-\theta) & \theta \end{pmatrix}.$$

where in equation (32), $\lambda_1(\theta) = \theta(1-\theta)^{-1}$, and the two distinct forms for the transition probabilities appearing in each row are

$$\{\theta(1-\theta)^{-1}\}^{K_{ij}}(1-\theta), \quad K_{ij} = 0, 1.$$

Again the values of the coefficients for θ vary from row to row, subject to condition (34) that their sum is always equal to 1, while those for $(1-\theta)$ also have a sum equal to 1.

REFERENCES

Bartlett, M. S. (1951). The frequency goodness of fit test for probability chains. *Proc. Camb. Phil. Soc.* **47**, 86–95.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press.

Frazer, R. A., Duncan, W. J. & Collar, A. R. (1947). *Elementary Matrices*. New York: Macmillan.

Fréchet, M. (1952). *Traité du Calcul des Probabilités et de ses Applications*, tome 2, fasc. III. Gauthier-Villars.

Huzurbazar, V. S. (1948). The likelihood equation, consistency and maxima of the likelihood function. *Ann. Eugen., Lond.*, **14**, 185–200.

Koopman, B. O. (1936). On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc.* **39**, 399–409.

Montroll, E. W. (1947). On the theory of Markoff chains. *Ann. Math. Statist.* **18**, 18–36.

Pitman, E. J. G. (1936). Sufficient statistics and intrinsic accuracy. *Proc. Camb. Phil. Soc.* **32**, 567–79.

Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New York: John Wiley.

SIGNIFICANCE TESTS FOR DISCRIMINANT FUNCTIONS AND LINEAR FUNCTIONAL RELATIONSHIPS

By E. J. WILLIAMS

Division of Mathematical Statistics, C.S.I.R.O., Melbourne

I. INTRODUCTION

In previous papers (Bartlett, 1951; Williams, 1952a) certain exact tests for the adequacy of a hypothetical discriminant function were derived. Later papers (Williams, 1952b, 1952c, 1953) showed how these tests could be applied in a number of situations of practical usefulness. The first object of the present paper is to extend the work in the above-mentioned papers and to show how the results obtained may be interpreted in terms of multiple linear regression. The calculations may indeed be carried out in the manner of a covariance analysis. The second object is to develop, along the same lines, exact tests for an assumed linear relationship among variables; this is a problem which has been discussed in various contexts by Koopmans (1937), Tintner (1945, 1946, 1950), Geary (1948, 1949), Bartlett (1948), Anderson (1951) and others. Since the question of determining underlying relationships has been given considerable attention in the literature from a number of different points of view, the opportunity is taken also to discuss and to attempt to unify the different approaches made—the use of information provided by instrumental variates, by grouping of the data, and by higher moments.

The reason for discussing discriminant functions and functional relationships in the same paper is because the two problems are really different aspects of the same problem. This has been well demonstrated by Geary (1948). If a single discriminant function is assumed adequate to describe differences among a number of p-variate populations, this assumption is equivalent to assuming that there exist $p-1$ linear relations among the means for the p variates; the means then lie on a line. In general, postulating that the differences among the populations are described by r discriminant functions is equivalent to postulating $p-r$ linear relationships (provided always that the number of populations considered is not less than p). The quantity r, the number of dimensions in which the population means lie, may be called the rank of the populations, and $p-r$ the degeneracy.

Thus the test for a single linear relationship is equivalent to the test for the adequacy of $p-1$ discriminant functions. In deriving significance tests for either a discriminant function or a linear relationship, the same principle is applied, though the function being tested has a different role in the two cases, and thus enters differently into the tests. In the simple bivariate case, the test for a linear relationship is exactly the same as the test for the discriminant function which is orthogonal to it.

The problems of this paper have been framed above in terms of an analysis of variance model, for testing the significance of differences between populations. This has been done in order to link them with those discussed in the earlier work (Williams, 1952a, b). A more general specification is in terms of a regression model, wherein the interrelationships between a set of p variates and another set of q variates are investigated. In such a model a discriminant function is better described as a canonical variate. Throughout the remainder of

this paper the problems will be posed in terms of the regression model but interpreted also in terms of analysis of variance. Though both specifications are formally equivalent it is easier sometimes to work with one and sometimes with the other.

Bartlett (1951) has shown that the different aspects of the adequacy of a discriminant function may be tested by means of a factorization of a certain determinantal ratio. A general calculus of such factorizations is developed, and it is shown how it can be applied in a number of the tests described in this paper.

The method of treatment used earlier (Williams 1952a) in deriving exact significance tests was to frame the questions asked of the data somewhat differently from those usually posed. A hypothetical discriminant function being proposed, the question is whether this function is concordant with the data. Since the hypothetical discriminant function leads to a sufficient statistic for the unknown population parameter, tests can be derived which are independent of the value of the parameter. These tests are the counterparts of those based on the latent roots (largest or smallest) of a matrix, but have the advantage of being exact, even in small samples. The same approach is adopted in this paper, in deriving tests both for discriminant functions and functional relationships.

The most important practical results given in the present paper are that the tests previously given for the adequacy of a hypothetical discriminant function, and the tests now given for a functional relationship, can be reduced to tests derived by a covariance analysis of the data. The computations are thereby simplified because it is not necessary to determine the latent roots of a matrix equation, the test functions being expressed either as ordinary adjusted sums of squares in a covariance analysis, or as the ratios of determinants of sums of squares and products. If, however, the 'best' discriminant function yielded by the sample (or the 'best' linear relationship) is to be evaluated, it is still necessary to evaluate the largest (smallest) latent root and the corresponding latent vector.

II. Notation, terminology and preliminary results

Since this paper covers a rather wide field it has been thought desirable to set out clearly the notation to be used throughout. While an attempt has been made to use notation in conformity with previous work on the subject, some departures have been made in the interests of clarity.

We consider two sets of variates,

$$X_i \quad (i = 1, 2, ..., p),$$
$$Y_j \quad (j = 1, 2, ..., q),$$

measured on each of $n + 1$ individuals. The X_i will be assumed to have a non-singular joint normal distribution, with linear regression on the Y_j. Formally, this is equivalent to $q + 1$ populations with the same p-variate normal distribution of values about the means in each, so that the variation in the sample may be analysed into the q degrees of freedom between groups and the $n - q$ within groups.

The discrimination problem is to determine from a set of data, either that linear function or set of linear functions of the X_i which have greatest correlation with the Y_j, or the adequacy of a set of given discriminant functions. On the other hand, the problem of determining functional relationships is to find the set of linear functions, if any, of the X_i which are uncorrelated with the Y_j, or to test a given set of such functions for their correlation with

24-2

the Y_j. Linear functions which are uncorrelated with the Y_j, either by hypothesis or in the sample, will be called null functions and their sample values null variates.

Although from the formal correlation point of view there is a symmetry between the X_i and the Y_j, the tests of significance developed here are always of a function (either a discriminant function or a null function) of the variables of one of the sets (in this paper the X_i), and so are not symmetrical in the two sets. In this respect the present approach differs from that based on the canonical correlations alone, in which the magnitudes of the canonical correlations are used to decide the number of discriminant functions (or in other words, the rank and degeneracy of the population correlations). Reiersøl (1945) emphasizes the distinction between the two sets of variates by designating the Y_j 'instrumental' variates to distinguish them from the 'investigational' variates X_i.

The canonical variates of the two sets will be denoted by x_i, y_i, and the corresponding squared canonical correlations by θ_i. The hypothetical discriminant function will be denoted by ξ, and its squared multiple correlation with the Y_j, or discriminant ratio, by λ.

The tests of significance for a discriminant function have been shown in earlier papers to depend only on the original discriminant ratios θ_i and the new discriminant ratios after the effect of ξ has been eliminated by covariance, which will be denoted by ϕ_i.

It will be shown similarly in this paper that the tests for a linear functional relationship, that is, for a given null function, depend on sets of original and new latent roots: the new latent roots which will be denoted by θ_i' are, however, defined differently from those used in discriminant analysis.

The following notation will be used:

Sums of squares and products:

t_{hi} sum of products of X_h and X_i
p_{ij} sum of products of X_i and Y_j
u_{jk} sum of products of Y_j and Y_k
b_{hi} sum of products of X_h and X_i for regression or between-group line of analysis of variance
w_{hi} sum of products of X_h and X_i for residuals from regresssion or within groups.

Also $t_{\xi i}$ sum of products of ξ and X_i
$t_{\xi\xi}$ sum of squares of ξ

and similarly for other sums of squares and products with the suffix ξ.

The corresponding matrices will be denoted by capitals. Then the following results are readily obtained, relating the different expressions occurring in the tests of significance:

$$B+W = T,$$
$$B = PU^{-1}P', \qquad |B|/|T| = \Pi\theta_i,$$
$$W = T - PU^{-1}P', \quad |W|/|T| = \Pi(1-\theta_i).$$

The matrix of sums of squares and products of all the $p+q$ variates will be denoted by S:

$$S = \begin{bmatrix} T & P \\ P' & U \end{bmatrix}.$$

Armed with these results, we are in a position to set out expressions in terms either of the regression model or the analysis of variance model.

In the derivation of tests of significance by means of covariance analysis use is made of sums of squares and products of the Λ after adjustment by certain covariance variates. Since there seems to be no consistent terminology for these quantities, we propose to use the term "adjusted" only for quantities which are adjusted without any loss of degrees of freedom and the term "reduced" for quantities for which the degrees of freedom are reduced by the number of covariance variates. Thus, the analysis of variance of any Λ would give the following partition of degrees of freedom:

	Degrees of freedom
Regression	q
Residual	$n - q$
Total	n

After adjustment by r covariance variates, the analysis would give

	Degrees of freedom
Regression, adjusted	q
Residual, reduced	$n - q - r$
Total, reduced	$n - r$

The adjusted regression sum of squares may, if required, be partitioned as follows:

	Degrees of freedom
Difference of regressions on covariance variates	r
Regression, reduced	$q - r$
Regression, adjusted	q

III. THE FACTORIZATION OF DETERMINANTAL RATIOS

The overall likelihood criterion for testing the rank of the population from which a sample has been drawn is, in terms of the sample latent roots, equal to

$$\prod_{i=1}^{p} (1 - \theta_i).$$

In terms of determinants of sums of squares and products it may also be expressed as

$$| W |/| T | = | T - PU^{-1}P' |/| T |$$
$$= \frac{|S|}{|T||U|},$$
(1)

and so may be represented as a determinantal ratio without calculation of the latent roots. The criterion may be expressed as the ratio of two determinants of order p, with $n-q$ and n degrees of freedom respectively; and since the X_i and the Y_j enter symmetrically into the expression, it may also be expressed as the ratio of two determinants of order q, with degrees of freedom $n-p$ and n. It will accordingly be denoted

$$(n; p, q)$$

to indicate its dependence on the three sets of degrees of freedom. In particular, for $p = 1$, the ratio of a residual sum of squares with $n-q$ degrees of freedom to a total sum of squares with n degrees of freedom is denoted $(n; 1, q)$.

Now if s is an integer less than q, a sum of squares with q degrees of freedom can be partitioned into two independent sums of squares with s and $q-s$ degrees of freedom. Corresponding to this partition, the ratio of sums of squares with $n-q$ and n degrees of freedom can be factorized into two independent factors: one with $n-s$ and n degrees of freedom, the other with $n-q$ and $n-s$ degrees of freedom; so that

$$(n; 1, q) = (n; 1, s)\,(n-s; 1, q-s).$$
(2)

In exactly the same way, determinantal ratios may be factorized, giving, for example,

$$(n; p, q) = (n; p, s)\,(n-s; p, q-s),$$
(3)

or, more generally, $= (n; p, s)\,(n-s; r, q-s)\,(n-r-s; p-r, q-s).$
(4)

These results are the basis of the factorizations of likelihood criteria given by Bartlett (1951), and will be applied repeatedly throughout this paper.

As in univariate analysis, the first factor in (3) is seen to be the likelihood criterion for the simple effect of the s variables, the remaining $q-s$ being ignored, while the second factor corresponds to the partial effect of the $q-s$ variables after the elimination of the first s. The partial factor is accordingly the one which is to be used in significance tests. Just as in univariate analysis, there are always two alternative factorizations, depending on which set out of the q variables is to be eliminated and which is to be the subject of test. Thus

$$(n; p, q) = (n; p, s)\,(n-s; p, q-s)$$
$$= (n-q+s; p, s)\,(n; p, q-s).$$
(5)

When $p = 1$, the tests for the partial factors may be simultaneously set out in the form of an analysis of variance:

Effect	Degrees of freedom	Sum of squares
Partial s variates	s	$1 - (n; 1, s)$
Partial $q-s$ variates	$q-s$	$1 - (n; 1, q-s)$
Residual	$n-q$	$(n; 1, q)$
Total	n	1

The possibility of thus arranging the tests arises from the fact that, for example,

$$\frac{1-(n;\,1,s)}{(n;\,1,q)} = \frac{1-(n-q+s;\,1,s)}{(n-q+s;\,1,s)}$$

$$= \frac{s}{n-q} F_{(s,\,n-q)} \tag{6}$$

The moments of the likelihood criterion may be determined by means of this factorization. Taking $s = 1$ and factorizing successively, we have

$$(n;\,p,q) = (n;\,p,1)(n-1;\,p,1)(n-2;\,p,1)\ldots(n-q+1;\,p,1).$$

It is readily shown that the tth moment about zero of $(n;\,p,1)$ is

$$E[(n;\,p,\,1)^t] = \frac{\Gamma[\frac{1}{2}(n-p)+t]\,\Gamma(\frac{1}{2}n)}{\Gamma[\frac{1}{2}(n-p)]\,\Gamma(\frac{1}{2}n+t)},$$

so that the moments of the likelihood criterion are given by

$$E[(n;\,p,q)^t] = \prod_{i=0}^{q-1} \frac{\Gamma[\frac{1}{2}(n-p-i)+t]\,\Gamma[\frac{1}{2}(n-i)]}{\Gamma[\frac{1}{2}(n-p-i)]\,\Gamma[\frac{1}{2}(n-i)+t]} \tag{7}$$

(cf. Bartlett (1938), equation (26)).

An important particular case is the limiting one when n tends to infinity. Writing

$$(p,q) = \lim_{n\to\infty} (n;\,p,q)^n,$$

we find that the factorization (3) becomes

$$(p,q) = (p,s)(p,q-s), \tag{8}$$

and it can be shown that $-\log(p,q)$ is distributed as a sum of squares with pq degrees of freedom (cf. Williams, 1952a, p. 26).

When n tends to infinity, the two factorizations given above are identical, and are equivalent to a simple partition of a sum of squares—as is, indeed, evident from the analysis set out above for the particular case $p = 1$.

An application. As a simple example of the application of these methods we consider the case where r of the population latent roots are known to be unity, and the vanishing of the remaining population roots is under test.

If r population roots are unity, it follows that $\theta_1, \theta_2, \ldots, \theta_r$ are all unity, and that the corresponding canonical variates $x_1, x_2, \ldots, x_r$ are equal (apart from possible changes of sign) to $y_1, y_2, \ldots, y_r$, the canonical variates of the second set. Therefore, in order to examine the residual variation of the x_i, the p variates of the first set may be replaced by any $p-r$ of them not linearly dependent on the first r of them. Each of these variates has $n-q$ residual degrees of freedom. For their regression on the y_i, the variates are restricted by the condition of orthogonality with the first $r\,x_i$ and hence with the first $r\,y_i$, so that the regression has but $q-r$ degrees of freedom. Accordingly, the residual likelihood criterion for testing the existence of the non-unit latent roots, namely,

$$\prod_{i=r+1}^{p} (1-\theta_i) \tag{9}$$

is a $(n-r; p-r, q-r)$ variable. In particular. if $r = p-1$ $(p \leqslant q)$, the single sample latent root θ_p may be tested by an F-test:

$$1 - \theta_p = (n-p+1; 1, q-p+1),$$

$$F = \frac{(n-q)\theta_p}{(q-p+1)(1-\theta_p)},\tag{10}$$

with $q-p+1$ and $n-q$ degrees of freedom.

Corresponding results hold when $q < p$.

These results have practical application where, even if the first r population canonical correlations are not unity, the sample latent roots are close to unity. Then the elimination of the corresponding sample canonical variates leads to a residual likelihood criterion (9) with approximately the distribution given. Such a result has already been discussed by Bartlett (1947) and others.

IV. Review and extension of previous work

Bartlett (1951) derived tests of significance for the goodness of fit of a single hypothetical discriminant function, in terms of its direction, and of departures of the data from col-linearity. These results generalized some results of Williams (1952a) for the particular case $p = 2$ (two variates) and $q = 2$ (two degrees of freedom between groups). However, two points were not clarified in this work. First, while Williams used for test functions the partial factors, Bartlett suggested that the use of partial factors was not necessary in these tests. It should be clear from the discussion in § III that the partial factors are appropriate for significance tests.

Secondly, it was stated by Bartlett that the general test functions could be expressed in terms of the original latent roots and the new latent roots after adjustment for the hypothetical discriminant function, but the expressions were not given. These results, which are given below, show the relationship of the exact tests to the approximate tests based on the individual ordered latent roots. Apart from this, they are of more analytic than practical interest, since the calculation of all the latent roots, before and after adjust-ment, is laborious. Moreover, in § V of this paper, simplified computational procedures based on the analysis of covariance will be given, which make possible the ready application of the exact tests without the calculation of latent roots.

Expression of test functions in terms of canonical correlations

For this section, we express all results in terms of the canonical variates x_i, y_i and the canonical correlations. Let the hypothetical discriminant function which is under test be

$$\xi = \Sigma m_i x_i, \quad \text{with} \quad \Sigma m_i^2 = 1,$$

so that the corresponding discriminant ratio is

$$\lambda = \Sigma m_i^2 \theta_i.$$

The first factorization of the overall likelihood criterion, representing the elimination of the effect of the hypothetical discriminant function, is

$$\prod_i (1-\theta_i) = (1-\lambda)\prod_j (1-\phi_j),$$

corresponding to $(n: p.q) = (n: 1.q)(n-1: p-1.q)$. and leading by the elimination of the factor $1 - \lambda$ for the hypothetical discriminant function to the residual likelihood criterion $\prod_j (1-\phi_j)$.

Bartlett (1951) gives two factorizations of the residual likelihood criterion providing tests for simple direction and partial collinearity (equation (7) of his paper) and for partial direction and simple collinearity (equation (20) of his paper). In our notation. these factorizations may be written

$$\prod_{\substack{j \\ (n-1;p-1,q)}} (1-\phi_j) = \prod_i (1-\theta_i)/(1-\lambda),$$

$$= \underbrace{\left[\frac{1-\sum m_i^2 \theta_i^2 \lambda}{1-\lambda}\right]}_{\substack{\text{Simple direction} \\ (n-1;p-1,1)}} \underbrace{\left[\frac{\prod(1-\theta_i)}{1-\sum m_i^2 \theta_i^2 \lambda}\right]}_{\substack{\text{Partial collinearity} \\ (n-2;p-1,q-1)}} \tag{11}$$

$$= \underbrace{\left[\frac{\lambda}{(1-\lambda)\sum \frac{m_i^2 \theta_i}{1-\theta_i}}\right]}_{\substack{\text{Partial direction} \\ (n-q;p-1,1)}} \underbrace{\left[\frac{\prod(1-\theta_i)}{\lambda}\sum \frac{m_i^2 \theta_i}{1-\theta_i}\right]}_{\substack{\text{Simple collinearity} \\ (n-1;p-1,q-1)}}. \tag{12}$$

The degrees of freedom for these factors agree with the values given by Bartlett. Since the first factor of (12) has unity for one of its degrees of freedom. it may be tested by an ordinary F-test.

It now remains to express the m_i^2 in terms of the original and new latent roots. From Williams (1952a, p. 20) we have. in the present notation. the equation for the ϕ_j

$$\sum_i \frac{m_i^2(1-\theta_i)}{\theta_i - \phi_j} = 0 \quad (j = 1, 2, ..., p-1). \tag{13}$$

On expressing this as an equation of degree $p-1$ in ϕ_j. we have

$$\phi_j^{p-1}\sum_i m_i^2(1-\theta_i) - \phi_j^{p-2}\sum_i m_i^2(1-\theta_i)(\sum_h \theta_h - \theta_i) + \ldots = 0.$$

i.e.

$$\phi_j^{p-1}(1-\lambda) - \phi_j^{p-2}[\sum_i \theta_i(1-\lambda) - \lambda + \sum_i m_i^2 \theta_i^2] + \ldots = 0.$$

Hence

$$\sum_j \phi_j = \sum_i \theta_i - (\lambda - \sum_i m_i^2 \theta_i^2)/(1-\lambda),$$

so that

$$(\lambda - \sum_i m_i^2 \theta_i^2)/(1-\lambda) = \sum_i \theta_i - \sum_j \phi_j,$$

and the factors (11) become

$$\left[\frac{\sum_i \theta_i - \sum_j \phi_j}{\lambda}\right]\left[\frac{\lambda \prod_j (1-\phi_j)}{\sum_i \theta_i - \sum_j \phi_j}\right].$$

The second factor provides the test for collinearity. It takes its maximum value $\prod_2^p (1-\theta_i)$, when $\lambda = \theta_1$. The use of this product therefore overestimates the significance of departures from collinearity. When $p = 2$ it reduces to

$$\frac{\frac{(1-\theta_1)(1-\theta_2)}{1-\lambda}}{\frac{(1-\theta_1)(1-\theta_2)}{1-\lambda} + \frac{\theta_1 \theta_2}{\lambda}} = [1 + v_2/(1-v_1)(1-v_2)]^{-1} \quad \text{(notation of Williams, 1952a),}$$

and is thus a function of the variance ratio previously given for testing collinearity.

When $q = 2$ the second factor of (11) becomes

$$\frac{(1-\theta_1)(1-\theta_2)}{\theta_1+\theta_2-\phi_1-\phi_2} = 1 - Q\,P \quad \text{(notation of Williams, 1952a)}$$

$$= [1+v_2/(1-v_1)(1-v_2)]^{-1},$$

which again corresponds to the test function given in Williams (1952a, p. 26).

For the alternative factorization it is convenient to write

$$\Theta = \theta/(1-\theta),$$

$$\Phi = \phi/(1-\phi),$$

$$\Lambda = \lambda/(1-\lambda).$$

Then equation (13) becomes

$$\sum_i \frac{m_i^2}{\Theta_i - \Phi_j} = 0 \quad (j = 1, 2, ..., p-1),$$

so that

$$\Phi_j^{p-1} - \Phi_j^{p-2} \sum_i m_i^2 (\sum_h \Theta_h - \Theta_i) + ... = 0.$$

Hence

$$\sum_i m_i^2 \Theta_i = \sum_i \Theta_i - \sum_j \Phi_j,$$

and the factors (12) become

$$\left[\frac{\Lambda}{\sum_i \Theta_i - \sum_j \Phi_j}\right] \left[\frac{\prod_j (1-\phi_j)}{\Lambda} (\sum_i \Theta_i - \sum_j \Phi_j)\right]. \tag{12'}$$

The first factor provides the test for direction; being a $(n-q;\, p-1,\, 1)$ variable it may be tested by means of an F-test with $p-1$ and $n-p-q+1$ degrees of freedom:

$$F = \frac{n-p-q+1}{p-1} \frac{(1-(n-q;\, p-1,\, 1))}{(n-q;\, p-1,\, 1)}.$$

When λ is close to θ_1, so that the hypothetical discriminant function differs little in direction from the first canonical variate, the first factor of (12) is approximately λ/θ_1. This provides a justification of the approximate F-test

$$F = \frac{n-p-q+1}{p-1} \frac{(\theta_1-\lambda)}{\lambda}.$$

When $p = 2$ the first factor becomes

$$\frac{(1-\theta_1)(1-\theta_2)}{(1-\lambda)(1-\theta_1\theta_2/\lambda)} = 1-v_1,$$

the test function previously given.

When $q = 2$ the first factor reduces to

$$(1-v_1)(1-v_2),$$

again equivalent to the test function given previously.

The two factorizations may be shown in closer relationship by means of the following representations, which make use of the fact that $(1-\phi_j) = (1+\Phi_j)$:

$$\prod_j (1-\phi_j) = \left[\frac{\sum_i \theta_i - \sum_j \phi_j}{\lambda}\right]\left[\frac{\lambda \prod_j (1-\phi_j)}{\sum_i \theta_i - \sum_j \phi_j}\right],$$

$$\prod_j (1+\Phi_j) = \left[\frac{\sum_i \Theta_i - \sum_j \Phi_j}{\Lambda}\right]\left[\frac{\Lambda \prod_j (1+\Phi_j)}{\sum_i \Theta_i - \sum_j \Phi_j}\right].$$

V. DERIVATION OF TESTS FOR DIRECTION AND COLLINEARITY
BY COVARIANCE ANALYSIS

It has been shown by Bartlett (1939) that a general test for the adequacy of a discriminant function is provided by the residual likelihood criterion after the elimination by covariance of the effect of the function. A point that has been overlooked, however, is that by a simple extension of this method the separate tests for direction and collinearity of the function may be derived. Bartlett there deals with the simple case of discrimination between two groups ($q = 1$). In the treatment below we shall consider the regression model, in the general case with q independent (or instrumental) variates.

(a) The case $p = 2$

To introduce the method, we consider first two variates X_1 and X_2, and their correlation with the q variates $Y_1, Y_2, ..., Y_q$. Let the hypothetical discriminant function be

$$\xi = l_1 X_1 + l_2 X_2.$$

To test the adequacy of ξ, we consider the residual correlation of the X_i with the Y_j after the elimination of ξ by covariance. We may adjust either X_1, X_2 or any linear function of them, for its dependence on ξ; the same analysis is achieved in any case, since ξ is itself a linear function of X_1 and X_2.

For the overall test, the sum of squares, with q degrees of freedom, for regression on the Y_j after adjustment for ξ, is tested against the reduced residual sum of squares, with $n - q - 1$ degrees of freedom. Now the adjusted regression sum of squares may be separated into two parts: (i) one with $q - 1$ degrees of freedom, namely, the sum of squares for regression, reduced by the elimination of the effect of ξ, and (ii) one with 1 degree of freedom, representing the regression of the regression values on ξ.

In the analysis of variance model with q degrees of freedom between groups, the two parts of the adjusted sum of squares between groups are (i) the reduced sum of squares between groups, and (ii) the sum of squares for the difference of between-group and within-group regressions.

Then it turns out that the reduced sum of squares (i) gives the sum of squares for testing the independence of the adjusted X_i and Y_j (or the collinearity of populations in the analysis of variance model), while the sum of squares (ii) provides the test for direction of the proposed discriminant function. The tests are equivalent to those given in the formal analysis of variance presented previously (Williams, 1952a, p. 24).

This equivalence may be established as follows.

With the notation given in §II, we can express the determinantal ratios in terms of symmetric functions of the latent roots (discriminant ratios). Thus

$$|B|/|T| = \theta_1\theta_2,$$

$$|W|/|T| = (1-\theta_1)(1-\theta_2),$$

while, by definition, $$b_{\xi\xi}/t_{\xi\xi} = \lambda.$$

Then the sums of squares for X_1, after adjustment by covariance for ξ, may be expressed in terms of determinantal ratios or latent roots as follows:

	Degrees of freedom	Sums of squares in terms of	
		Determinantal ratios	Latent roots
Difference of regressions (direction)	1	$l_2^2(|T|/t_{\xi\xi} - |B|/b_{\xi\xi} - |W|/w_{\xi\xi})$	$\dfrac{(\theta_1-\lambda)(\lambda-\theta_2)}{\lambda(1-\lambda)}\dfrac{l_2^2|T|}{t_{\xi\xi}}$
Between-groups reduced (collinearity)	$q-1$	$l_2^2|B|/b_{\xi\xi}$	$\dfrac{\theta_1\theta_1}{\lambda}\dfrac{l_2^2|T|}{t_{\xi\xi}}$
Within-groups reduced	$n-q-1$	$l_2^2|W|/w_{\xi\xi}$	$\dfrac{(1-\theta_1)(1-\theta_2)}{1-\lambda}\dfrac{l_2^2|T|}{t_{\xi\xi}}$
Total reduced	$n-1$	$l_2^2|T|/t_{\xi\xi}$	$\dfrac{l_2^2|T|}{t_{\xi\xi}}$

The right-hand column gives terms which are proportional to the quantities used in testing direction and collinearity against a residual term. This analysis further justifies the procedure given in the earlier papers, of testing the partial factors (i.e. terms corresponding to the terms given in the above covariance analysis) rather than the simple factors.

This approach to the tests for the adequacy of a hypothetical discriminant function greatly simplifies the computations. By this method it is not necessary to calculate the θ_i, nor even to determine the canonical variates x_i. A straightforward covariance analysis is all that is required. However, if what is required is not a test of the adequacy of a given discriminant function but the calculation of the most satisfactory discriminant function from the data, then the canonical analysis is still necessary.

(b) The case $p > 2$

The procedure outlined above may be generalized for greater values of p. We take the hypothetical discriminant function to be

$$\xi = l_1X_1 + l_2X_2 + \ldots + l_pX_p,$$

and consider the analysis of covariance of the X_i on ξ. Since ξ is a linear function of the X_i, the number of linearly independent variates after the elimination of ξ is reduced to $p-1$. As for the case $p = 2$, the sums of squares and products of the residual variates may be

separated into terms for direction and collinearity, the term for direction having a single degree of freedom. From the $p-1$ residual variates a linear function may be made up in the same way as a discriminant function for a comparison for which the sum of squares for direction is relatively maximized. This sum of squares clearly has $p-1$ degrees of freedom. When the sum of squares is maximized relative to the reduced total sum of squares, we get the sum of squares for the simple direction effect (collinearity being ignored) (11). When it is maximized relative to the reduced sum of squares for residuals from regression on the Y_j, we get the sum of squares for the partial direction effect (12). The analysis will be demonstrated for the simple criterion for direction; the partial criterion may be similarly derived.

For any two of the variates X_h and X_i, the sum of products for direction may be derived by difference in the manner given above (when $p = 2$), but may be expressed more simply for the present purpose as

$$d_h d_i b_{\xi\xi} w_{\xi\xi} t_{\xi\xi},$$

where

$$d_i = \left(\frac{b_{\xi i}}{b_{\xi\xi}} - \frac{w_{\xi i}}{w_{\xi\xi}}\right).$$

Now consider some linear compound of the X_i, say

$$\zeta = l_1' X_1 + l_2' X_2 + \ldots + l_p' X_p.$$

For this compound, the sum of squares for direction is

$$\left(\sum_{i=1}^{p} l_i' d_i\right)^2 b_{\xi\xi} w_{\xi\xi}/t_{\xi\xi},$$

while the total sum of squares is

$$\sum_1^p \sum_1^p l_h' l_i' t_{hi}.$$

To maximize the ratio of direction to total, which we denote by D, with respect to the l_i', we have

$$\sum_1^p l_h' t_{hi} D t_{\xi\xi} = d_i \sum_1^p l_h' d_h b_{\xi\xi} w_{\xi\xi}.$$

The scale of the l_i' being arbitrary, we may so choose it that

$$D t_{\xi\xi} = \sum_1^p l_h' d_h b_{\xi\xi} w_{\xi\xi}, \tag{14}$$

whence

$$\sum_1^p l_h' t_{hi} = d_i. \tag{15}$$

If we write the typical element of the inverse of T as t^{hi}, then the solutions of (15) are given by

$$l_i' = \sum_1^p t^{hi} d_h.$$

Hence, from (14),

$$D = \sum_1^p \sum_1^p t^{hi} d_h d_i b_{\xi\xi} w_{\xi\xi}/t_{\xi\xi}.$$

It is to be noted that ζ is orthogonal to ξ. This may be obvious from its derivation; alternatively, the total sum of products of ζ and ξ is

$$\sum_1^p \sum_1^p l_h l_i' t_{hi} = \sum_1^p l_h d_h$$

$$= 0,$$

while the sum of products for direction is proportional to

$$\sum_1^p \sum_1^p l_h l_i' d_h d_i = 0.$$

The quantity D, being an invariant, may readily be expressed in terms of latent roots in order to show the equivalence of the present results with those given in §IV. Transforming to the canonical variates x_i, and putting as before

$$\xi = \Sigma m_i x_i,$$

we have

$$d_i = \frac{m_i(\theta_i - \lambda)}{\lambda(1-\lambda)},$$

$$D = \frac{\Sigma m_i^2(\theta_i - \lambda)^2}{\lambda(1-\lambda)}$$

$$= \frac{\Sigma m_i^2 \theta_i^2 - \lambda^2}{\lambda(1-\lambda)},$$

so that

$$1 - D = \frac{1 - \Sigma m_i^2 \theta_i^2 / \lambda}{1-\lambda},$$

which agrees with (11).

To derive the test function for partial collinearity in terms of determinantal ratios, we make use of the factorization (11).

The residual likelihood criterion

$$\Pi(1 - \theta_i)/(1 - \lambda) = \frac{|W|}{|T|} \frac{t_{\xi\xi}}{w_{\xi\xi}}.$$

Hence the partial collinearity factor is

$$\frac{|W|}{|T|} \frac{t_{\xi\xi}}{w_{\xi\xi}} \Big/ (1 - D) = \frac{|W|}{|T|} \frac{t_{\xi\xi}^2}{w_{\xi\xi}} \Big/ \left(t_{\xi\xi} - b_{\xi\xi} w_{\xi\xi} \sum_1^p \sum_1^p t^{hi} d_h d_i \right), \tag{16}$$

which is a $(n-2; p-1, q-1)$ variate.

In the same way, the partial direction factor may be found by maximizing the ratio of the sum of squares for direction to the reduced residual sum of squares. The maximized ratio is

$$\frac{t_{\xi\xi}}{t_{\xi\xi} + b_{\xi\xi} w_{\xi\xi} \Sigma w^{hi} d_h d_i}. \tag{17}$$

This is a $(n-q; p-1, 1)$ variate, so may be tested by the F-test:

$$F = \frac{n-p-q+1}{p-1} \frac{b_{\xi\xi} w_{\xi\xi}}{t_{\xi\xi}} \Sigma w^{hi} d_h d_i,$$

with $p-1$ and $n-p-q+1$ degrees of freedom.

It may be verified that these results agree with the general results given by Bartlett (1951) and the particular results given by Williams (1952a) for $p = 2, q = 2$.

VI. GENERAL REMARKS ON THE LINEAR FUNCTIONAL RELATIONSHIP

As has often been remarked, the linear functional relationship between two variates subject to error is different from either of the regression relationships between the two: the relationships are identical only when the independent variate in the regression relationship is

errorless. Also the two relationships have different applications. The regression relationship is the more generally useful, relating as it does to observed values; an important application is the prediction of either observed or 'true' values of one variate from observed values of the other. The functional relationship connects the 'true' values of the two variates; its use is in theoretical studies of underlying laws. In most practical applications the regression and not the functional relationship is required. Since we are here considering only linear relationships, the word 'linear' will usually be omitted.

The regression relationships are based on the variation in both the 'true' values and the random errors to which they are subject, the functional relationship on the variation in the 'true' values alone. A little reflexion will show (and examination of the literature will confirm, see, for example, Haavelmo (1943)) that the functional relationship is therefore relevant only to a study of how the 'true' values of both variates are affected by some extraneous variate or variates; that is to say, the relationship shows what elements of the system are invariant under changes in conditions. It is not of interest to know the underlying relationship (if any) between two variates when each is affected only by random error; usually what is then wanted is one or other of the regression relationships.

It is well known that, when both the 'true' values and the errors are normally distributed, the functional relationship cannot be determined from data. Lindley (1947), Reiersøl (1950) and others have proved a number of theorems which show that the non-normality of one of the distributions is necessary for the estimation of the relationship to be possible. Thus it may be taken that functional relationships are not determinable from the internal analysis of a set of variates with normal distributions. Put in another way, since the first- and second-order sample cumulants summarize all the information in samples from normal populations, it follows that information beyond the first- and second-order cumulants of the distributions must be available if the functional relationship is to be determinable. This information may be present in the sample, if the distributions are not normal, for then the first- and second-order sample cumulants are no longer sufficient statistics; alternatively, it may be introduced through knowledge of the relationship of each variate with some extraneous variates. As mentioned above, it is only under such circumstances, when some such additional information exists, that the functional relationship is of interest; in other words, whenever functional relationships are of practical interest they are also determinable from data.

In passing it may be mentioned that Berkson's (1950) method of 'controlled' variables, as elucidated by Lindley (1953), is really a means of dealing with one variate as though it were errorless, and bringing the estimation of the functional relationship back to the estimation of a regression equation.

The foregoing discussion could be generalized for the determination of relationships among three or more variates, for which the same general considerations apply. We shall be concerned henceforth with the determination of functional relationships among a set of variates through their relationships with the variates of a second set. Reiersøl (1941, 1945) has termed the second set *instrumental* variates, to distinguish them from the first set of *investigational* variates. This approach has been used also by Geary (1943, 1949) and others.

Another method (Wald, 1940; Bartlett, 1949) for determining the functional relationship uses groupings of the values of the variates. Neyman & Scott (1951) have shown that only 'in very exceptional circumstances' does the method provide consistent estimates. Roughly

speaking, the method leads to consistent results provided the separation of values of one of the variates is sufficiently wide in the neighbourhood of the group limits, so that the grouping based on observed values is equivalent to a grouping based on 'true' values. Clearly, under these conditions, the differences between groups may be attributed to some extraneous variates, so that the method of grouping is then roughly equivalent to the method of instrumental variates.

Both methods have this in common, that they introduce extraneous variation affecting all the variables between which the relationship is found. The existence or assumption of this variation is basic to the methods.

In the next sections of this paper we give some exact tests for the existence of and for the constants in linear functional relationships, based on the use of instrumental variates. We shall assume for convenience that the investigational variates are subject to normally distributed errors, but make no particular assumptions about the distribution of the instrumental variates. Indeed, the instrumental variates may even be comparisons among groups or populations, in accordance with the general specification outlined in the Introduction.

VII. Tests for a single functional relationship

(a) Preliminary remarks

In § VI it was pointed out that consistent estimates of a functional relationship between two variates could be determined provided there was one instrumental variate. In general, a functional relationship among p variates can be determined if there are $p-1$ or more instrumental variates. In this section we shall consider, not the determination of the functional relationship, but tests for the concordance with the data of a given relationship. This approach parallels that adopted above in the sections on discriminant functions. It is possible to test the adequacy of the given relationship even when the data do not provide enough information to enable consistent estimates to be determined.

We shall consider, following the set-up given in the earlier sections, that there are p investigational variables X_i and q instrumental variables Y_j. The given functional relationship among the X_i will be defined by a null function ξ. It will be realized that, when p exceeds q, it is possible to choose $p-q$ independent linear functions of the original X_i whose correlation with the Y_j (or sum of squares between groups) vanishes. These functions are equivalent to the $p-q$ canonical variates corresponding to the identically vanishing latent roots. Thus it is not possible to determine from the sample a unique null function.

On the other hand, the test for a given null function is always possible. The general test for the adequacy of the function can be considered in two parts: first, a test for the conformity of the function in direction, with $p-1$ degrees of freedom; and secondly, a test for the residual correlation (or reduced variation between groups), with $q-p+1$ degrees of freedom. When $q < p$, no test for residual correlation is possible; the test for direction then has but q degrees of freedom.

(b) One instrumental variate ($q = 1$)

In order to gain some insight into the problems involved in testing the adequacy of a specified null function, we consider first the simplest case, where there is but one instrumental variate, Y. In this case, the test is one of the direction of the chosen null function. It may be most simply viewed in the following way. The null variate ξ has a sum of squares

for regression on Y which bears to the total sum of squares a ratio λ. It is required to test whether this is large enough to indicate real association. The discriminant function x based on the sample gives a ratio θ, leaving a residual $1 - \theta$, with $n - p$ degrees of freedom. Hence the efficient test, in which the residual has been minimized with respect to all the X_i, is given by the following analysis:

	Degrees of freedom	Sum of squares
Additional accounted for by discriminant function x	$p-1$	$\theta - \lambda$
Direction of null variate	1	λ
Residual	$n-p$	$1 - \theta$
Total	n	1

so that
$$F_{(1, n-p)} = \frac{(n-p)\lambda}{1-\theta}.$$

In terms of determinantal ratios, we have
$$\lambda = b_{\xi\xi}/t_{\xi\xi}, \quad 1-\theta = |W|/|T|,$$
so that
$$F_{(1, n-p)} = (n-p)\frac{b_{\xi\xi}}{t_{\xi\xi}}\frac{|T|}{|W|}.$$

This latter representation would give the simplest way of making the test.

(c) More than one instrumental variate ($q \geq 2$)

The method given above for $q = 1$ cannot be directly adapted to the more general case. The approach that is adopted is as follows. By hypothesis, the null variate ξ is uncorrelated with the instrumental variates, so that its total variation is homogeneous. In any sample, it is possible to choose a set of $p-1$ linearly independent functions of the X_i which are uncorrelated with ξ; these may be taken to define the direction of ξ in the p-space. We shall denote these $p-1$ functions by X'_i ($i = 1, 2, ..., p-1$).

Now consider the analysis of any one of the original X_i, or of any linear function of them not linearly dependent on the X'_i. We can carry out an analysis of covariance of this variate with the $p-1$ variates X'_i, to determine the reduced sums of squares for regression on the Y_j, for residual variation and for total variation (or between groups, within groups and total in the analysis of variance model). Then, provided $q \geq p$, we have the following effects appearing:

Effect	Degrees of freedom
Difference of regressions	$p-1$
Regression on the Y_j, reduced	$q-p+1$
Residual, reduced	$n-p-q+1$
Total, reduced	$n-p+1$

This analysis tests the existence of any correlation of the X_i with the Y_j, apart from that accounted for by the X'_i, and hence tests the correlation of the proposed null variate. The second term tests the residual correlation of ξ with the Y_j and the first term the direction; each may be tested against the reduced residual variance.

The analysis and significance tests are equivalent, whichever X-variate (subject to the above-mentioned conditions) be chosen, since the elimination of the X'_i by covariance effectively reduces the number of residual variates to one. In practice it is often simplest to take the X-variate as ξ itself. The fact that the adjustment of ξ by covariance leaves the total sum of squares unaltered does not affect the results; the degrees of freedom for the total line are still reduced to $n-p+1$.

It may be verified that, when $p = 2$, the analysis is exactly the same as that giving the test of significance for the discriminant function X'_1. This is as it should be, since the hypothesis that X'_1 is the population discriminant function is equivalent to the hypothesis that the variate orthogonal to it (i.e. ξ) is a null variate.

The analysis has its parallel in the tests developed for this purpose from Fisher's (1938) results by Tintner (1945, 1946), and based on the values of the latent roots. For the existence of a single null function (or linear functional relationship), Tintner takes as his test function the smallest of p latent roots. As has been shown by Hsu (1941), this smallest latent root is distributed approximately as a sum of squares with $q-p+1$ degrees of freedom. This approximation becomes exact in the limit when all the other latent roots are near to unity (apparently even if only the sample latent roots approach unity, regardless of the values of the population roots). This result has been established above (10) as a particular result on the factorization of determinantal ratios, except that there the term for 'difference of regressions' has not been segregated from the residual. In the derivation of the tests to be given below, in terms of canonical correlations, the relationship of the present tests with the approximate tests will be made clear.

(i) *Analysis in terms of canonical correlations*

The results of the analysis of covariance may be expressed in terms of the squared canonical correlation coefficients or discriminant ratios $\theta_1, \theta_2, ..., \theta_p$, between the sets of canonical variates x_i and y_i. Let the hypothetical null-variate be

$$\xi = \Sigma m_i x_i, \quad \Sigma m_i^2 = 1.$$

Then the set of $p-1$ x-variates orthogonal to ξ may be taken as

$$x'_i = x_i - m_i \xi.$$

Now of the y_i, p have non-zero correlation with the corresponding x_i, and $q-p$ have zero correlation with any x_i. They may be transformed to another set of variates in the following way: (1) the set of $p-1$ variates most closely correlated with the x'_i; (2) a variate η, uncorrelated with the x'_i but correlated with the x_i, and (3), as before, the $q-p$ variates uncorrelated with the x_i.

It is clear that the variate η represents the reduced correlation of ξ with the y_i. It is readily defined by the conditions given; for let

$$\eta = \Sigma n_i y_i;$$

then
$$\Sigma \eta x'_h = 0 \quad (h = 1, 2, ..., p-1).$$

Hence

$$\Sigma \eta x_\lambda = m_\lambda \Sigma \xi \eta,$$

or

$$n_\lambda \theta_\lambda^{\frac{1}{2}} = m_\lambda \Sigma m_i n_i \theta_i^{\frac{1}{2}},$$

so that

$$n_\lambda \propto m_\lambda \theta_\lambda^{-\frac{1}{2}},$$

and, since

$$\Sigma n_\lambda^2 = 1,$$

$$n_\lambda = m_\lambda \theta_\lambda^{-\frac{1}{2}} (\Sigma m_i^2/\theta_i)^{-\frac{1}{2}}.$$

For the regression of ξ on η, the sum of squares, with $q - p + 1$ degrees of freedom, is

$$(\Sigma \xi \eta)^2 = n_\lambda^2 \theta_\lambda / m_\lambda^2 = (\Sigma m_i^2/\theta_i)^{-1}. \tag{18}$$

This is the sum of squares for regression reduced by elimination of the x_i'. Clearly it vanishes if $q < p$ (for then $\theta_p = 0$), and in general takes values between θ_p and θ_1.

In the same way it may be shown that the reduced residual sum of squares (i.e. the sum of squares of residuals of ξ from regression on the y_i and the x_j'), with $n - p - q + 1$ degrees of freedom, is

$$(\Sigma m_i^2/(1 - \theta_i))^{-1}. \tag{19}$$

Finally, by subtraction, the sum of squares for difference of regressions (i.e. for the test of direction of the proposed null-variate), with $p - 1$ degrees of freedom, is

$$1 - (\Sigma m_i^2/\theta_i)^{-1} - (\Sigma m_i^2/(1 - \theta_i))^{-1}. \tag{20}$$

These sums of squares may also be expressed in terms of the latent roots θ_i and the latent roots for the variates x_i' and y_i, which will be denoted by θ_i' $(i = 1, 2, \ldots, p-1)$. Since ξ is orthogonal to all the x_i', it follows that the product of the latent roots θ_i', multiplied by the sum of squares for regression of ξ on η, is equal to the product of the original latent roots; that is,

$$\Pi \theta_i' (\Sigma m_i^2/\theta_i)^{-1} = \Pi \theta_i.$$

Similarly,

$$\Pi (1 - \theta_i') (\Sigma m_i^2/(1 - \theta_i))^{-1} = \Pi (1 - \theta_i).$$

The analysis of variance may then be set out as follows:

	Degrees of freedom	Sum of squares
Direction	$p-1$	$1 - \Pi \theta_i \cdot \Pi \theta_i' - \Pi(1 - \theta_i) \cdot \Pi(1 - \theta_i')$
Correlation	$q-p+1$	$\Pi \theta_i / \Pi \theta_i'$
Residual	$n-p-q+1$	$\Pi(1 - \theta_i)/\Pi(1 - \theta_i')$
Total	$n-p+1$	1

If the latent roots θ_i and the direction cosines m_i are known, the simplest procedure is to calculate the sums of squares from the expressions (18), (19) and (20).

The sum of squares for correlation always exceeds θ_p, but approaches θ_p when the direction of ξ approaches that of x_p. Hence it is seen that the use of θ_p as a sum of squares with $q - p + 1$ degrees of freedom, according to the approximate tests, will underestimate the significance of the correlation.

If $q < p$, the analysis reduces to

	Degrees of freedom	Sum of squares
Direction	q	$1 - \Pi(1-\theta_i)/\Pi(1-\theta'_i)$
Residual	$n-p-q+1$	$\Pi(1-\theta_i)/\Pi(1-\theta'_i)$
Total	$n-p+1$	1

there being in this case no test for correlation.

When $q = 1$ this analysis reduces to that previously given: for it may be shown that, in general,

$$\lambda = \Sigma\theta_i - \Sigma\theta'_i,$$

so that, in this case $(1-\theta_1)/(1-\theta'_1) = 1-\lambda/(1-\theta_1+\lambda)$

and the sums of squares for direction and residual are respectively

$$\lambda/(1-\theta_1+\lambda)$$

and

$$(1-\theta_1)/(1-\theta_1+\lambda),$$

which are proportional to those previously given.

It may readily be verified that, when $p = 2$, these sums of squares reduce to those given for a test of a single discriminant function.

(ii) *Analysis in terms of determinantal ratios*

For computing purposes it is more straightforward, when the individual latent roots and canonical variates are not required, to express the results just given in terms of determinantal ratios.

We consider the analysis of the null variate ξ ($= \Sigma l_i X_i$, no longer necessarily normalized so that its sum of squares is unity), deriving the reduced sum of squares for regression, with $q-p+1$ degrees of freedom, and the reduced residual sum of squares, with $n-p-q+1$ degrees of freedom. The sum of squares for 'difference of regressions', with $p-1$ degrees of freedom, is obtained by subtraction of these two items from the (reduced) total.

We may take the X'_i to be

$$X'_i = X_i - \xi t_{\xi i}/t_{\xi\xi} \quad (i = 1, 2, ..., p-1),$$

and shall write

P_i for the vector $(p_{i1}, p_{i2}, ..., p_{iq})$,
T_i for the vector $(t_{i1}, t_{i2}, ..., t_{ip})$,
T_{i-} for the vector $(t_{i1}, t_{i2}, ..., t_{i,p-1})$,
P for the matrix consisting of the first $p-1$ rows of P.
and T for the matrix consisting of the first $p-1$ rows and columns of T.

To calculate the reduced sum of squares for regression, the sums of squares and products of the X'_i and ξ for the regression of each on the Y_j are first determined. For example, the sum of products of X'_i with Y_j is

$$p_{ij} - p_{\xi j}t_{\xi i}/t_{\xi\xi},$$

so that the regression sum of products of X'_h and X'_1 is

$$(P_h - P_\xi t_{\xi h}/t_{\xi\xi})\, U^{-1}(P'_1 - P'_\xi\, t_{\xi 1}/t_{\xi\xi}).$$

Accordingly, the matrix of regression sums of squares and products of the X'_1 is

$$(P_- - T'_{\xi-}P_\xi/t_{\xi\xi})\, U^{-1}(P'_- - P'_\xi T_{\xi-}/t_{\xi\xi}). \tag{21}$$

Now the reduced sum of squares for the regression of ξ on the Y_j is the ratio of two determinants: that of the regression sums of squares and products of the X'_1 and ξ, and that of the regression sums of squares and products of the X'_1 alone. By transformation without change of scale, the X'_1 and ξ may be replaced by $X_1, X_2, \ldots, X_{p-1}$, and $l_p X_p$, hence the determinant in the numerator is equal to

$$l_p^2\,|PU^{-1}P'| = l_p^2\,|B|.$$

The required denominator is the determinant of the matrix (21), which is found, by elementary transformations, to be

$$-\frac{l_p^2}{t_{\xi\xi}^2}\begin{vmatrix} 0 & T_\xi \\ T'_\xi & PU^{-1}P' \end{vmatrix} = -\frac{l_p^2}{t_{\xi\xi}^2}\begin{vmatrix} 0 & T \\ T' & B \end{vmatrix}$$

$$= \frac{l_p^2}{t_{\xi\xi}^2}\,|B|\,T_\xi B^{-1}T'_\xi.$$

Hence the reduced sum of squares for regression is

$$t_{\xi\xi}^2/T_\xi B^{-1}T'_\xi.$$

In the same way, it may be shown that the reduced residual sum of squares is

$$t_{\xi\xi}^2/T_\xi W^{-1}T'_\xi.$$

Also, since
$$T_\xi T^{-1}T'_\xi = t_{\xi\xi},$$

we see that the reduced total sum of squares, which is given by the analogous formula

$$t_{\xi\xi}^2/T_\xi T^{-1}T'_\xi$$

is equal to $t_{\xi\xi}$, as it should be.

On elimination of common factors from each term, the analysis of variance may be set out as follows:

	Degrees of freedom	Sum of squares
Direction	$p-1$	(by subtraction)
Regression, reduced	$q-p+1$	$(T_\xi B^{-1}T'_\xi)^{-1}$
Residual, reduced	$n-p-q+1$	$(T_\xi W^{-1}T'_\xi)^{-1}$
Total, reduced	$n-p+1$	$(T_\xi T^{-1}T'_\xi)^{-1} = t_{\xi\xi}^{-1}$

It may readily be verified that, when the variates are transformed to the canonical set, the terms in this analysis of variance reduce to those given in the previous subsection.

VIII. GENERALIZATIONS

It has been shown above that the test for a given functional relationship as developed in this paper is really equivalent to the test for the adequacy of $p-1$ discriminant functions to represent the variation in the data. Thus we have treated here only the tests for 1 or $p-1$ discriminant functions. Tests for any number r of discriminant functions may be derived in the same way, but the cases treated above are the simplest to deal with, since they each deal with only one function (discriminant function or null function); fiducial limits for the constants in these functions can therefore be derived using these significance tests.

If r hypothetical discriminant functions are given, their effect may be eliminated, giving the following factorization of the likelihood criterion:

$$(n; p, q) = (n; r, q)(n-r; p-r, q).$$

The second factor on the right is the residual likelihood criterion, which may be further factorized to give criteria for direction and collinearity:

$$(n-r; p-r, q) = (n-r; p-r, r) \qquad (n-2r; p-r, q-r)$$
$$\text{simple direction} \qquad\qquad \text{partial collinearity}$$

$$= (n-q; p-r, r) \qquad (n-r; p-r, q-r)$$
$$\text{partial direction} \qquad\qquad \text{simple collinearity}$$

When $r = p-1$, one of the degrees of freedom in each factor is unity, which is the reason why the tests for a single null function can be thrown into the form of an analysis of variance. We then have

$$(n-p+1; 1, q) = (n-p+1; 1, p-1) \quad (n-2p+2; 1, q-p+1)$$
$$= (n-q; 1, p-1) \qquad (n-p+1; 1, q-p+1)$$

which corresponds to the analysis of variance set-up previously given.

There seem to be practical applications of tests for more than one functional relationship; for example, Bartlett (1948) discusses the problem of estimating supply and demand relationships. On the other hand, it seems that seldom if ever is more than one discriminant function required in specifying differences among populations; if these differences cannot be interpreted in terms of one discriminant function, what is then required is some likelihood test as a means of classifying the observations into the different populations.

REFERENCES

ANDERSON, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Statist.* 22, 327.

BARTLETT, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.* 34, 33.

BARTLETT, M. S. (1939). A note on tests of significance in multivariate analysis. *Proc. Camb. Phil. Soc.* 35, 180.

BARTLETT, M. S. (1947). Multivariate analysis. *J. R. Statist. Soc. Supp.* 9, 176.

BARTLETT, M. S. (1948). A note on the statistical estimation of supply and demand relations from time series. *Econometrica*, 16, 323.

BARTLETT, M. S. (1949). Fitting a straight line when both variables are subject to error. *Biometrics*, 5, 207.

BARTLETT, M. S. (1951). The goodness of fit of a single hypothetical discriminant function in the case of several groups. *Ann. Eugen., Lond.*, 16, 199.

BERKSON, J. (1950). Are there two regressions? *J. Amer. Statist. Ass.* 45, 164.

FISHER, R. A. (1938). The statistical utilization of multiple measurements. *Ann. Eugen., Lond.*, 8, 376.

GEARY, R. C. (1943). Relations between statistics: the general and the sampling problem when the samples are large. *Proc. R. Irish Acad.* A, 49, 177.

GEARY, R. C. (1948). Studies in relations between economic time series. *J. R. Statist. Soc.* B, 10, 140.

GEARY, R. C. (1949). Determination of linear relations between systematic parts of variables with errors of observation the variances of which are unknown. *Econometrica*, 17, 30.

HAAVELMO, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11, 1.

HSU, P. L. (1941). On the problem of rank and the limiting distribution of Fisher's test function. *Ann. Eugen., Lond.*, 11, 39.

KOOPMANS, T. (1937). *Linear Regression Analysis of Economic Time Series*. Haarlem: DeErven, F., Bohn, N.V.

LINDLEY, D. V. (1947). Regression lines and the linear functional relationship. *J. R. Statist. Soc.* Supp. 9, 218.

LINDLEY, D. V. (1953). Estimation of a functional relationship. *Biometrika*, 40, 47.

NEYMAN, J. & SCOTT, ELIZABETH L. (1951). On certain methods of estimating the linear structural relation. *Ann. Math. Statist.* 22, 352.

REIERSOL, O. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, 9, 1.

REIERSOL, O. (1945). Confluence analysis by means of instrumental sets of variables. *Ark. Mat. Astr. Fys.* A, 32, no. 4.

REIERSOL, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica*, 18, 375.

TINTNER, G. (1945). A note on rank, multicollinearity and multiple regression. *Ann. Math. Statist.* 16, 304.

TINTNER, G. (1946). Multiple regression for systems of equations. *Econometrica*, 14, 5.

TINTNER, G. (1950). A test for linear relations between weighted regression coefficients. *J. R. Statist. Soc.* B, 12, 273.

WALD, A. (1940). The fitting of straight lines if both variables are subject to error. *Ann. Math. Statist.* 11, 284.

WILLIAMS, E. J. (1952a). Some exact tests in multivariate analysis. *Biometrika*, 39, 17.

WILLIAMS, E. J. (1952b). The interpretation of interactions in factorial experiments. *Biometrika*, 39, 65.

WILLIAMS, E. J. (1952c). Use of scores for the analysis of association in contingency tables. *Biometrika*, 39, 274.

WILLIAMS, E. J. (1953). Tests of significance for concurrent regression lines. *Biometrika*, 40, 297.

THE USE OF TRANSFORMATIONS AND MAXIMUM LIKELIHOOD IN THE ANALYSIS OF QUANTAL EXPERIMENTS INVOLVING TWO TREATMENTS

By F. YATES, F.R.S.

Rothamsted Experimental Station

If the effect of a treatment is of the quantal (all or nothing) type, e.g. it kills or does not kill an insect, the results of experiments will be in the form of proportions of experimental units affected. If t such treatments are compared in a single experiment in which the experimental units are randomly allocated, the differences between the observed proportions will give estimates of the differences between the treatments, and a χ^2 test on the associated $2 \times t$ table ($t-1$ d.f.) provides an overall test of significance. If, however, in order to increase the precision, the units are subdivided before allocation into groups which are relatively homogeneous within themselves, the experiment will be analogous to an ordinary randomized block experiment, the groups being the blocks, but the quantities requiring analysis (corresponding to the 'plot yields') will be proportions instead of quantitative measurements. When a set of experiments is carried out, with random allocation within each experiment, but with possible lack of homogeneity between experiments, the situation is similar, the groups being in this case experiments.

If the numbers of experimental units in the various cells of the group × treatment or experiment × treatment table are approximately equal, or more generally if they are proportionate, the analysis presents no great difficulty. The use of the angular transformation, and an analysis of variance of the transformed values, will frequently be all that is required. Indeed, if the variation in susceptibility from experiment to experiment is not large, the results will not be seriously distorted if the data from the different groups or experiments are pooled and treated as homogeneous. If, however, the numbers in the different cells vary in an irregular manner, as is often the case in material of this kind, pooling is inadmissible and inequality in the weights prevents straightforward analysis.

The application of methods of estimation to quantal data has been extensively developed in recent years through the use of transformations in conjunction with maximum likelihood. Exact analytical methods were first developed for the probit transformation, in connexion with biological assay, but similar methods have recently been developed for other transformations of which the logit and log log are of most interest in this connexion. These methods are beginning to be used for the analysis of quantal data of many kinds (see, for example, Jolly, 1950; Dyke & Patterson, 1952; and Yates, 1953, § 9·7), but their utility for dealing with experimental data of the type we are considering is still insufficiently realized. Combination of probabilities, for example, is still commonly used to provide a test of significance (sometimes called the probability integral test) for sets of experiments involving two treatments.

The estimation approach automatically provides a test of significance as a by-product of the estimate and its estimate of error. Its greatest advantage, however, is that if an effect is demonstrated we are immediately provided with an estimate of its magnitude in

meaningful terms, together with approximate fiducial limits. At the same time answers are provided to certain ancillary questions regarding the homogeneity of the data.

The analysis of a group of experiments containing several treatments has been discussed by Jolly using the probit transformation. The present paper deals in more detail with the case of two treatments, particularly in respect of errors and tests of significance. The appropriateness of the various transformations under different circumstances is also considered in more detail than by Jolly. A further new point is the use of weighted estimates to give a first approximation. The investigation originated in a request by Prof. Gert Bonnier of Stockholm for advice on the exact procedure for applying the combination of probabilities test to a set of genetical experiments on mutation rates in *Drosophila*, and I am much indebted to him for permission to reproduce the data here. The numbers involved in these experiments were sufficiently large for the data to be analysed by large-sample methods, and this analysis will first be described, as it brings out many of the points at issue without the complexities associated with the use of transformations and the method of maximum likelihood. The maximum-likelihood method will then be developed, and applied to this set of data and to three examples given by Pearson (1950).

THE COMBINATION OF PROBABILITIES TEST

If we have a set of k experiments (or observational data) comparing two treatments, the results of each experiment can be set out in the form of a 2×2 table. A test of significance on the data as a whole can be made by calculating the significance level P for each experiment separately, and then combining these probabilities by the method first suggested by Fisher in the fourth edition of *Statistical Methods for Research Workers* (1932), and independently by Karl Pearson (1933). This consists of calculating the sum of the quantities $-2\log_e P$, which are the values of χ^2 for 2 d.f. corresponding to the significance levels P. If there is no difference between the treatments, then, subject to certain qualifications, $S(-2\log_e P)$ will be distributed as χ^2 with $2k$ d.f.

This test and the analogous test provided by direct summation of the values of χ^2 for the separate experiments have been discussed by various authors, in particular Cochran (1942, 1952), Lancaster (1949) and Pearson (1950). Certain further points emerged in the course of the present investigation, but these are best dealt with in a separate paper (Yates, 1955). Here it is only necessary to make the following general points:

(1) Application of the combination of probabilities test does not provide an analysis which is in any sense complete, since no estimate of the magnitude of the treatment difference is provided.

(2) The test must be in some degree inefficient. There are two main reasons for this:

(a) No account is taken of the relative accuracy of the different experiments when combining the results.

(b) $S(-2\log_e P)$ does not provide an efficient estimate of the treatment difference, even when all experiments are of the same accuracy and the difference is small, and consequently cannot be the basis of an efficient test of significance.

It is perhaps also worth noting here that Fisher did not envisage the use of the combination of probabilities test on data of the type we are considering. He put it forward for use in cases in which it is desired to obtain a single test of the significance of an aggregate of probabilities, 'taking account only of these probabilities, and not of the detailed composition of the data from which they are derived, which may be of very different kinds'.

DROSOPHILA DATA: THE LARGE SAMPLE APPROACH

The data with which we are concerned are shown in Table 1. They were derived from a set of six experiments on the frequency of lethals in paternal X-chromosomes from irradiated spermatozoa of *Drosophila melanogaster*. The experiments were carried out to test whether there was any difference in this frequency when the irradiation was given to spermatozoa which were harboured (F) in the females' receptacles, and (M) in the males' testicles. In the first four experiments a dose of 960 r. was given, and in the last two a dose of 3000 r. In addition to the change of irradiation rate it was believed that unavoidable environmental changes between experiments might influence the mutation rates. An account of the experiments has now been published (Bonnier & Lüning, 1953).

Table 1. *Frequency of mutants with two methods of irradiation*

Exp.	Method F			Method M		
	Mutant	Normal	% mutant	Mutant	Normal	% mutant
1	29	815	3·44	45	1486	2·94
2	78	2622	2·89	27	1438	1·84
3	110	3175	3·35	100	4281	2·28
4	25	1038	2·35	52	2053	2·47
5	31	339	8·38	43	507	7·82
6	57	543	9·50	27	436	5·83

It will be seen that in all experiments except Exp. 4 method F gives a higher mutation rate than method M. We require to test the significance of this difference, and obtain an estimate of its magnitude.

In order to develop a suitable estimation procedure we must specify in mathematical terms the quantities that require estimation. This specification will depend on the phenomena that are being investigated. In the present case we may postulate that with a given method of irradiation and an experiment of standard sensitivity a unit of irradiation has a given small chance θ' of producing a mutation at a given locus. The chance that λ units of irradiation will produce a mutation at that locus will then be

$$1 - e^{-\lambda\theta'}.$$

If several loci are involved, with probabilities θ', θ'', etc., and $\theta' + \theta'' + \ldots = \theta$, the total probability of one or more mutations will be

$$1 - e^{-\lambda(\theta'+\theta''+\cdots)} = 1 - e^{-\lambda\theta}.$$

If the conditions vary from experiment to experiment a sensitivity factor μ_r will have to be introduced, giving a total probability of mutation for experiment r of $1 - e^{-\lambda_r\mu_r\theta}$, where λ_r is the irradiation rate for the rth experiment.

With two methods of irradiation the probabilities of mutation in experiment r can therefore be written as

$$\pi_{1r} = 1 - \exp[-\lambda_r\mu_r\theta_1],$$
$$\pi_{2r} = 1 - \exp[-\lambda_r\mu_r\theta_2].$$

Since μ_r is unknown it must be eliminated, giving

$$\frac{\theta_1}{\theta_2} = \frac{\log(1-\pi_{1r})}{\log(1-\pi_{2r})}.$$

Substituting the observed mutation rates p_{1r} and p_{2r} for π_{1r} and π_{2r} we obtain an estimate of θ_1/θ_2 from the rth experiment. These separate estimates can then be combined by weighting inversely as their estimated variances.

In the present example the mutation rates are sufficiently small for π_{1r} and π_{2r} to be taken as approximately equal to $\lambda_r \mu_r \theta_1$ and $\lambda_r \mu_r \theta_2$ respectively. There will be certain advantages in working with estimates of

$$\theta_1 - \theta_2$$
$$\tfrac{1}{2}(\theta_1 + \theta_2)$$

instead of θ_1, θ_2, i.e. the ratio of the difference of the mutation rates for experiments of unit sensitivity per unit irradiation to the mean of these mutation rates. The quantity

$$E_r = \frac{p_{1r} - p_{2r}}{\tfrac{1}{2}(p_{1r} + p_{2r})}$$

will provide an estimate of this ratio. An approximation to the variance of E_r is easily calculated by large-sample theory as follows:

$$\frac{\partial E_r}{\partial p_{1r}} = \frac{4p_{2r}}{(p_{1r}+p_{2r})^2}, \quad \frac{\partial E_r}{\partial p_{2r}} = \frac{-4p_{1r}}{(p_{1r}+p_{2r})^2}.$$

If the rth contingency table is written

	F	M	Total
Mutant	n'_{1r}	n'_{2r}	n'_r
Normal	$n_{1r}-n'_{1r}$	$n_{2r}-n'_{2r}$	$n_r-n'_r$
Total	n_{1r}	n_{2r}	n_r

so that $p_{1r} = n'_{1r}/n_{1r}$, $p_{2r} = n'_{2r}/n_{2r}$, we have

$$V(p_{1r}) = \frac{\pi_{1r}(1-\pi_{1r})}{n_{1r}}, \quad V(p_{2r}) = \frac{\pi_{2r}(1-\pi_{2r})}{n_{2r}}, \quad \operatorname{cov} p_{1r}p_{2r} = 0.$$

Hence, approximately,

$$V(E_r) = \frac{16 p_{1r} p_{2r}}{(p_{1r}+p_{2r})^4}\left[\frac{p_{2r}q_{1r}}{n_{1r}} + \frac{p_{1r}q_{2r}}{n_{2r}}\right]. \tag{1}$$

This formula is approximate both because we are dealing with a non-linear function and because π_{1r} and π_{2r} have been replaced by p_{1r} and p_{2r}. If we are concerned with testing the significance of the difference of p_{1r} and p_{2r} we may replace them both by the pooled estimate $p_r = n'_r/n_r$ from the marginal totals of the table. Since the factor $1/(p_{1r}+p_{2r})$ occurs in E_r the factor $1/(p_{1r}+p_{2r})^2$ may be left unchanged in the variance. We then have

$$V(E_r) = \frac{4}{(p_{1r}+p_{2r})^2}\frac{p_r q_r n_r}{n_{1r} n_{2r}}. \tag{2}$$

It may be noted that under these circumstances

$$\frac{E_r^2}{V(E_r)} = \frac{n_r\{n_{1r}(n_{2r}-n'_{2r})-n_{2r}(n_{1r}-n'_{1r})\}^2}{n_{1r}n_{2r}n'_r(n_r-n'_r)},$$

which is the ordinary expression for χ^2 in a 2×2 contingency table, without correction for continuity.

As has been shown by Cochran (1942) and others, the correction for continuity is not required when combining data from several tables. If, therefore, we take a weighted mean of the estimates E_r with weights $1/V(E_r)$, the resultant test of significance may be expected to be adequately accurate when the values of χ^2 corrected for continuity give satisfactory approximations to the one-tail probabilities of the distributions generated by the 2×2 tables. This will be the case in the present instance, since all expectations are reasonably large and n_{1r} and n_{2r} differ by less than a factor of 2 in all experiments, and are very nearly equal in total.

The calculations for $\bar{E}$ using formula (2) for $V(E_r)$ are shown in Table 2. The weighted mean of E is

$$\bar{E} = 44\cdot182/153\cdot85 = 0\cdot2872,$$

with standard error $1/\sqrt{153\cdot85} = 0\cdot0806$. The level of significance (one tail) is therefore $P = 0\cdot000184$.

Table 2. *Calculation of $\bar{E}$*

Exp.	E	V(E)	w	wE
1	+0·1559	0·05459	18·32	2·856
2	+0·4421	0·04622	21·63	9·563
3	+0·3786	0·01790	55·88	21·156
4	−0·0489	0·05776	17·31	−0·846
5	+0·0692	0·05099	19·61	1·357
6	+0·4785	0·04739	21·10	10·096
	+0·2872 ± 0·0806	—	153·85	44·182

$$P = 0\cdot000184.$$

The results of Table 2 also provide a simple test of the constancy of $(\theta_1 - \theta_2)/\frac{1}{2}(\theta_1+\theta_2)$ over the different experiments. All that is necessary is to calculate the weighted sum of squares of deviations of E. This is

$$Sw_r(E_r-\bar{E})^2 = Sw_r E_r^2 - \bar{E}Sw_r E_r$$

$$= 4\cdot962.$$

If there is no real variation this will be distributed as χ^2 with 5 d.f. This gives $P = 0\cdot42$, so there is no evidence of variation.

The same procedure may be adopted to test whether there is any evidence of variation in the mutation rates from experiment to experiment, apart from those due to differences in irradiation rate. For this purpose it will be appropriate to test whether the estimates of

$\frac{1}{2}(\theta_1 - \theta_2)$ are homogeneous. These estimates are given in Table 3, together with their standard errors calculated in the conventional manner. The value of the weighted sum of squares of deviations of these rates is 5·32 (χ^2 with 5 d.f.) giving $P = 0.38$. There is thus no evidence of any such variation. The postulated law also accounts satisfactorily for the difference in mutation rates at the different levels of irradiation.

Table 3. *Estimates of* $\frac{1}{2}(\theta_1 + \theta_2) \times 10^4$

Exp.	
1	0·332 ± 0·0398
2	0·246 ± 0·0248
3	0·293 ± 0·0201
4	0·251 ± 0·0300
5	0·270 ± 0·0306
6	0·256 ± 0·0270
Weighted mean	0·272 ± 0·0110

It may be noted that if the irradiation rate had been constant, and the difference between θ_1 and θ_2 could be neglected, the ordinary χ^2 test on the 2×6 contingency table of the values of n'_r and $n_r - n'_r$ could be used to test the constancy of the mutation rate from experiment to experiment.

Since the difference between the p_{1r} and p_{2r} is demonstrated we might revise the calculations by using formula (1) for $V(E_r)$. This, however, is merely a partial step in the direction of the small sample solution by maximum likelihood, which will be outlined in the following sections.

Cochran (1954) has adopted a similar procedure to that given above in order to provide a test of significance for a set of 2×2 contingency tables. His test criterion is the weighted mean of the quantities $p_{1r} - p_{2r}$. This is, in fact, equivalent to a weighted mean of the quantities

$$E'_r = \frac{p_{1r} - p_{2r}}{p_r q_r}.$$

E'_r differs from E_r only in the substitution of the pooled estimate p_r and the introduction of the factor q_r, which gives a difference, analogous to logits, which may be expected to be reasonably constant over a wide range of p_r. Regarded in this way, therefore, Cochran's analysis could easily be extended to provide a test for the constancy of the difference from experiment to experiment.

MAXIMUM-LIKELIHOOD SOLUTION: CHOICE OF TRANSFORMATION

The maximum-likelihood solution is most easily obtained by the use of a suitable transformation, as, for example, the probit transformation, which has become familiar in toxicology and biological assay. The maximum-likelihood solution can then be obtained by successive approximation. The choice of transformation depends on the mathematical specification (the 'model') which is considered most appropriate to the data under

investigation. When this specification has been decided the transformation chosen should be such that the quantities that require estimation are simple functions of the transformed variate; thus in probit work it is anticipated that the transformed variate will bear a linear relation to the dosage when the latter is expressed in suitable units (usually logarithms), and the quantities that require estimation are the constant term and the slope of this regression. A secondary requirement is that in no part of the range (or at least in no part covered by the data) do the working values or the weights associated with the transformed variate become infinite.

In the genetical example given above the transformation

$$y = \log(-\log q)$$

is obviously suitable. For with this transformation if Y_{1r} and Y_{2r} represent the transformed values of π_{1r} and π_{2r} we have

$$Y_{1r} = \log \lambda_r + \log \mu_r + \log \theta_1,$$

$$Y_{2r} = \log \lambda_r + \log \mu_r + \log \theta_2.$$

The effects of the treatments and of the variations in sensitivity are therefore represented by additive components in the transformed scale. This transformation, which is the same as the ordinary log log transformation with p replaced by q, may be termed the *complementary log log transformation*.

In many cases in which there is no clearly appropriate theoretical model, and in which there is no reason to differentiate between the two ends of the probability scale, the logit transformation

$$z = \tfrac{1}{2} \log(p/q)$$

is likely to be suitable. When the p's are all small the logit transformation is equivalent to one-half the complementary log log transformation, and when the p's are all nearly unity the logit transformation is equivalent to one-half the ordinary log log transformation.*

If the logit transformation is used the analysis will be conducted on the assumption that the treatment and sensitivity effects are additive on the transformed scale, i.e. that, except for random variations, the difference in the logits for the two treatments is constant from experiment to experiment.

In certain circumstances the probit transformation may be considered to be more appropriate than the logit transformation. Except at the extreme ends of the scale, however, there is little difference between them, and in the absence of any theoretical reasons for choosing one or the other it will rarely be possible to determine, by reference to the data themselves, which specification gives a better representation of the effects under investigation.

MAXIMUM-LIKELIHOOD EQUATIONS

As has been demonstrated in a number of places, the general procedure of obtaining the maximum-likelihood solution when using a transformation of the form $y = \phi(p)$, is as follows:

(1) Determine a set of N provisional values Y of the transformed variate for the N observed proportions. These can conveniently be deduced from preliminary (here called first) estimates of the required parameters or, when regression lines are involved, graphically.

---

* In the author's *Sampling Methods* (2nd edition, 1953) the logit transformation is defined as $z = \log p/q$. This does not accord with modern usage, and was due to an over-hasty reference to Finney's *Statistical Method in Biological Assay* (1952).

In many cases it pays, when making the first estimate to use weights similar to those used in the further steps of the analysis.

(2) For each observed proportion p (given by say n successes out of n observations) calculate a working value given (when y increases uniformly with p) by

$$y = pY_{max.} + qY_{min.}$$
$$= Y_{max.} - q(\text{Range})$$
$$= Y_{min.} + p(\text{Range}).$$

where $Y_{max.}$ and $Y_{min.}$ are the maximum and minimum working values corresponding to the assigned provisional value Y, and $Y_{max.} - Y_{min.} = \text{Range}$. Determine also the weight W to be assigned to the working value. This is given by nw, where w is the weighting coefficient. $Y_{max.}$, $Y_{min.}$ and w are functions of Y (or of the corresponding untransformed proportion P) whose form depends on the transformation. The general formulae are

$$Y_{max.} = Y - Q\left(\frac{dY}{dP}\right), \quad Y_{min.} = Y - P\left(\frac{dY}{dP}\right), \quad \text{Range} = \left(\frac{dY}{dP}\right), \quad w = \frac{1}{PQ}\bigg/\left(\frac{dY}{dP}\right)^2.$$

For the transformations commonly in use tables of these functions are readily available.

If y decreases uniformly with increasing p the formulae for $Y_{max.}$ and $Y_{min.}$ are interchanged and p and q are interchanged in the formulae for the working value y.

(3) Using the working values as observed variates obtain revised estimates of the parameters by the ordinary method of least squares, weighting each observation by the weight determined as above.

(4) From the revised estimates of the parameters determine a new set of provisional values, and repeat the whole process as often as is necessary.

In the case of the genetic example already discussed in large sample terms we may write

$$\log \theta_1 = t_1, \quad \log \theta_2 = t_2,$$
$$\log \lambda_r = l_r, \quad \log \mu_r = m_r \quad (r = 1, ..., k).$$

t_1, t_2 and the m_r are the parameters that have to be estimated.

If t_1', t_2' and m_r' are first estimates of these parameters provisional values are given by the equations

$$\left.\begin{array}{l} Y_{1r} = l_r + m_r' + t_1', \\ Y_{2r} = l_r + m_r' + t_2'. \end{array}\right| \tag{3}$$

Experiment r gives the two observation equations

$$t_1 + m_r = y_{1r} - l_r + \epsilon_{1r} \text{ weight } W_{1r},$$
$$t_2 + m_r = y_{2r} - l_r + \epsilon_{2r} \text{ weight } W_{2r},$$

there being k such pairs of equations in all.

The ordinary method of least squares then gives the normal equations:

$$t_1 \Sigma W_{1r} + \Sigma m_r W_{1r} = \Sigma W_{1r}(y_{1r} - l_r),$$
$$t_2 \Sigma W_{2r} + \Sigma m_r W_{2r} = \Sigma W_{2r}(y_{2r} - l_r),$$
$$m_r(W_{1r} + W_{2r}) + t_1 W_{1r} + t_2 W_{2r} = W_{1r}(y_{1r} - l_r) + W_{2r}(y_{2r} - l_r) \quad (r = 1, ..., k),$$

where Σ denotes summation over r from 1 to k and t_1, t_2 and the m_r now denote estimates. These equations are of the same form as equation (1) of Jolly's paper.

Since one of these equations is redundant we may conveniently put

$$\Sigma m_r(W_{1r} + W_{2r}) = 0. \tag{4}$$

The sum of the first two equations then gives

$$t_1 \Sigma W_{1r} + t_2 \Sigma W_{2r} = \Sigma W_{1r} y_{1r} + \Sigma W_{2r} y_{2r} - \Sigma(W_{1r} + W_{2r}) l_r. \tag{5}$$

Substituting the values of m_r given by the last k equations in either of the first two equations, we obtain

$$t_1 - t_2 = \left\{ \Sigma \frac{W_{1r} W_{2r}}{W_{1r} + W_{2r}} (y_{1r} - y_{2r}) \right\} \Big/ \Sigma \frac{W_{1r} W_{2r}}{W_{1r} + W_{2r}}. \tag{6}$$

Thus $t_1 - t_2$ is a weighted mean of $y_{1r} - y_{2r}$, with weights $W_{1r} W_{2r}/(W_{1r} + W_{2r})$.

To obtain m_r equations (5) and (6) may be solved for t_1 and t_2. We then have

$$m_r + l_r = \frac{1}{W_{1r} + W_{2r}} \{W_{1r}(y_{1r} - t_1) + W_{2r}(y_{2r} - t_2)\}. \tag{7}$$

When the logit transformation is used and the treatment and sensitivity effects are assumed additive on the logit scale, the solution is identical with that given above, except for the omission of the l terms.

APPLICATION TO DROSOPHILA DATA

The computations leading to the normal equations are set out in Table 4. The first estimates have been obtained by taking values of y corresponding to the observed p, and weighting coefficients corresponding to these y. Thus from Table XVI of Finney (1952) the y value corresponding to $1 - 0.0344$ is -3.37, and from Table XVII the corresponding weighting coefficient is 0.033, so that $W = 0.033 \times 844 = 28$. The values of $y_1 - y_2$, $W_1 + W_2$ and $W_1 W_2/(W_1 + W_2)$ are then calculated. These last are used as weights in calculating the weighted mean of $y_1 - y_2$, namely, $+0.2990$. The equations for t_1' and t_2' are therefore

$$332 t_1' + 290 t_2' = -6505 \cdot 33,$$

$$t_1' - t_2' = +0 \cdot 2990.$$

The numerical term of the first equation is obtained from the sum of products of the y and W columns less those of the l and $W_1 + W_2$ columns.

Solution in the ordinary manner gives

$$t_1' = -10 \cdot 3193, \quad t_2' = -10 \cdot 6183.$$

Equations (7) then give the values of $m_r' + l_r$ shown in Table 4.

Provisional values Y_{1r} and Y_{2r} are now calculated from equations (3). Thus

$$Y_{11} = -10 \cdot 3193 + 7 \cdot 0357 = -3 \cdot 28.$$

The working values and weights are calculated as explained above, using the appropriate table (Table XVII of Finney). It may be noted that if, as here, interpolation is required it is easiest to calculate two values of y and interpolate between these.

The remainder of the calculations proceed exactly as before, and lead to the second estimates shown in Table 4.

Table 4. *Analysis of Drosophila data*

Exp.	Basic data		1st estimates		2nd estimates		
	n_1	$100p_1$	y_1	W_1	Y_1	y_1	W_1
1	844	3·44	-3·37	28	-3·28	-3·349	31·2
2	2700	2·89	-3·53	81	-3·47	-3·528	75·0
3	3285	3·35	-3·37	108	-3·41	-3·378	108·8
4	1063	2·35	-3·72	26	-3·49	-3·711	31·9
5	370	8·38	-2·43	32	-2·30	-2·428	35·3
6	600	9·50	-2·30	57	-2·37	-2·301	53·6
				332			333·8

Exp.	n_2	$100p_2$	y_2	W_2	Y_2	y_2	W_2
1	1531	2·94	-3·53	46	-3·58	-3·509	42·1
2	1465	1·84	-4·01	26	-3·87	-3·979	30·3
3	4381	2·28	-3·76	97	-3·71	-3·767	106·0
4	2105	2·47	-3·68	51	-3·79	-3·683	47·0
5	550	7·82	-2·51	43	-2·60	-2·504	39·4
6	463	5·83	-2·82	27	-2·67	-2·803	31·0
				290			295·8

Exp.	l	y_1-y_2	W_1+W_2	$\dfrac{W_1W_2}{W_1+W_2}$	m'_r+l_r	y_1-y_2	W_1+W_2	$\dfrac{W_1W_2}{W_1+W_2}$	m_r+l_r	m_r
1	6·8669	+0·16	74	17·4	7·0357	+0·160	73·3	17·92	7·0543	+0·1874
2	6·8669	+0·48	107	19·7	6·6853	+0·451	105·3	21·58	6·7532	-0·1137
3	6·8669	+0·39	205	51·1	6·9062	+0·389	212·8	53·20	6·9010	+0·0341
4	6·8669	-0·04	77	17·2	6·8238	-0·028	78·9	19·00	6·8071	-0·0598
5	8·0064	+0·08	75	18·3	8·0149	+0·076	74·7	18·62	8·0134	+0·0070
6	8·0064	+0·52	84	18·3	7·9483	+0·502	84·6	19·64	7·9493	-0·0571
		+0·2990	622	142·0		+0·2937	629·6	149·96		

t_1	-10·3193		-10·3265
t_2	-10·6183		-10·6202

From the small changes in the values of y_1 and y_2 and the still smaller change (from +0·2990 to +0·2937) in t_1-t_2, which is trivial compared with the range of the values of which it is the weighted mean, it is clear that no further approximation is required. (A further cycle actually gives a value of +0·29341 for t_1-t_2.)

The advantage of deriving the first estimates by weighting the directly transformed proportions will now be apparent. The computations follow the same routine in each case (except for the calculation of working values) and a very satisfactory first approximation is

26

obtained. Moreover, the similarity of the figures in the two sets of computations provides a useful check against gross errors.*

The maximum-likelihood solution may be compared with the large-sample solution already obtained. For comparative purposes a correction to E is required to allow for the fact that squares of $\lambda\theta_1$ and $\lambda\theta_2$ have been neglected in the large-sample solution. This is obtained by multiplying each component of E by the appropriate value of $[1 - \frac{1}{2}(p_1 + p_2)]$. With this adjustment, and the corresponding adjustment in the weights, the large-sample estimate becomes $0\cdot2930$, with weight $147\cdot8$. The maximum-likelihood solution gives $\theta_1/\theta_2 = 1\cdot3414$, so that $(\theta_1 - \theta_2)/\frac{1}{2}(\theta_1 + \theta_2) = 0\cdot2916$. The agreement between the two methods of estimation is thus very close, the discrepancy being less than $1/50$ of the standard error.

ESTIMATION OF ERROR AND TESTS OF SIGNIFICANCE

Three types of test will commonly be required:

(1) A test of the difference between the two treatments, i.e. a test of whether $t_1 - t_2$ is significantly different from zero. Whether or not the difference is significant the estimated standard error of $t_1 - t_2$ should be given so that approximate fiducial limits can be assigned.

(2) A test whether there is any real variation in sensitivity of experiments, i.e. whether the differences between the m_r are significant.

(3) A test whether the residual variation is greater than can be accounted for by binomial sampling.

In the analysis of normally distributed data with variances in known ratios, i.e. with known relative weights, any set of parameters of which estimates have been obtained by the method of least squares can be tested for significance by finding the difference in the sum of squares accounted for when all parameters are included in the specification and that accounted for when the parameters under test are omitted from the specification. This difference is tested against the residual sum of squares by means of the z distribution. For the corresponding χ^2 test on quantal material, however, the relevant χ^2 cannot be exactly determined in this way, since changes in specification result in changes in the expectations, which in turn alter the expected variances.

A simple example will illustrate this point. Suppose we have two sets of $\frac{1}{2}n$ observations (n even), which are known to be derived from Poisson distributions whose means μ_1 and μ_2 are possibly different. If $\mu_1 = \mu_2$ the value of χ^2, χ_T^2 say, based on the whole of the data will be approximately distributed as χ^2 with $n - 1$ d.f. Likewise the value of χ^2, χ_D^2 say, calculated from the difference of the sums of the two sets of observations, will be distributed as χ^2 with 1 d.f. Whether or not $\mu_1 = \mu_2$ variation of the observations within sets will give two values of χ^2, χ_1^2 and χ_2^2 say, each of which will be distributed as χ^2 with $\frac{1}{2}n - 1$ d.f. It can easily be shown that

$$\chi_T^2 = \chi_D^2 + \frac{\bar{x}_1}{\bar{x}}\chi_1^2 + \frac{\bar{x}_2}{\bar{x}}\chi_2^2$$

$$= \chi_D^2 + \chi_1^2 + \chi_2^2 + \frac{\bar{x}_1 - \bar{x}_2}{\bar{x}}(\chi_1^2 - \chi_2^2),$$

where $\bar{x}_1$ and $\bar{x}_2$ are the observed means of the two sets and $\bar{x} = \frac{1}{2}(\bar{x}_1 + \bar{x}_2)$. If instead of χ_D^2, the difference between the total and residual χ^2, i.e. $\chi_T^2 - \chi_1^2 - \chi_2^2$, is taken as the criterion

* In general we may expect to save at least a cycle of iteration by initial weighting. In Dyke and Patterson's example, for instance, the first estimates obtained in *Sampling Methods* by weighting are fully as accurate as Dyke & Patterson's second estimates.

for testing the hypothesis $\mu_1 = \mu_2$ the discrepancy given by the last term of the above equation will be obtained.

It is therefore advisable to use direct tests for each set of parameters instead of relying on differences between χ^2 values. The general rule for obtaining the variances of the estimates of parameters in maximum likelihood solutions of this type is to invert the matrix of the coefficients of the final set of normal equations, thus obtaining a c matrix of the customary type. The variances and covariances are then given directly by the c values.

In the present case the fact that one of the constants is redundant introduces a slight additional complication. The procedure has been given by Yates & Hale (1939). Suppose that the h normal equations $E_1 = 0$, $E_2 = 0$, ..., $E_h = 0$ are governed by the identical relationship

$$\lambda_1 E_1 + \lambda_2 E_2 + ... + \lambda_h E_h = 0, \tag{8}$$

and that the relation

$$\mu_1 b_1 + \mu_2 b_2 + ... + \mu_h b_h = 0 \tag{9}$$

is assumed between the estimates of the parameters. Then instead of replacing the numerical terms of the normal equations by 1, 0, ..., 0 in order to obtain the equations for $c_{11}, c_{12}, ..., c_{1h}$. they must be replaced by

$$1 - \frac{\lambda_1 \mu_1}{S(\lambda \mu)}, \quad -\frac{\lambda_1 \mu_2}{S(\lambda \mu)}, \quad ...., \quad -\frac{\lambda_1 \mu_h}{S(\lambda \mu)}.$$

Similarly the equations for $c_{21}, c_{22}, ..., c_{2h}$ will be obtained by replacing the numerical terms by

$$-\frac{\lambda_2 \mu_1}{S(\lambda \mu)}, \quad 1 - \frac{\lambda_2 \mu_2}{S(\lambda \mu)}, \quad ...., \quad -\frac{\lambda_2 \mu_h}{S(\lambda \mu)},$$

etc.*

Designating the normal equations of our example by the suffices a, b, 1, 2, ..., k, we have

$$-E_a - E_b + E_1 + E_2 + ... + E_k = 0,$$

so that

$$\lambda_a = \lambda_b = -1, \quad \lambda_1 = \lambda_2 = ... = \lambda_k = 1.$$

Also from (4)

$$\mu_a = \mu_b = 0, \quad \mu_r = W_{1r} + W_{2r} \quad (r = 1, ..., k).$$

Consequently $S(\lambda \mu) = \Sigma(W_{1r} + W_{2r})$. The correction terms to be added to each of the first two sets of auxiliary equations are therefore

$$0, \quad 0, \quad g_1, \quad g_2, \quad ...., \quad g_k,$$

where $g_r = (W_{1r} + W_{2r}) / \Sigma(W_{1r} + W_{2r})$. For the remaining k sets the same correction terms are subtracted.

Tests for the t's

The c's can now be easily evaluated. All that is necessary is to make the appropriate changes in the numerical terms of the normal equations and utilize the solutions already obtained.

* The process can be extended to two (or more) redundant constants by introducing additional correction terms $-\mu_1' \lambda_1' S(\mu' \lambda')$, etc., provided that all the sums of products of the type $S(\lambda \mu')$, $S(\lambda' \mu)$, etc., are zero. The Yates & Hale paper is incorrect in that it omits this condition, and also assigns a numerical term μ_0 instead of zero to equation (9).

Putting $U = \Sigma \dfrac{W_{1r} W_{2r}}{W_{1r} + W_{2r}}$, we find

$$c_{aa} = V(t_1) = \frac{1}{\Sigma(W_{1r} + W_{2r})} + \frac{(\Sigma W'_{2r})^2}{U\{\Sigma(W_{1r} + W_{2r})\}^2},$$

$$c_{ab} = \operatorname{cov}(t_1 t_2) = \frac{1}{\Sigma(W_{1r} + W_{2r})} - \frac{\Sigma W'_{1r} \Sigma W'_{2r}}{U\{\Sigma(W_{1r} + W_{2r})\}^2}.$$

Hence

$$V(t_1 - t_2) = \frac{1}{U} = 1 \bigg/ \Sigma \frac{W_{1r} W_{2r}}{W_{1r} + W_{2r}},$$

as could, indeed, be deduced directly from the form of the expression for $t_1 - t_2$, and

$$V\tfrac{1}{2}(t_1 + t_2) = \frac{1}{\Sigma(W_{1r} + W_{2r})} + \frac{\{\Sigma(W_{1r} - W_{2r})\}^2}{4U\{\Sigma W_{1r} + W_{2r}\}^2}.$$

Test for the m's

Using a similar procedure for set r, and putting

$$T_r = \frac{1}{U} \frac{W_{1r} \Sigma W_{2r} - W_{2r} \Sigma W_{1r}}{(W_{1r} + W_{2r}) \Sigma(W_{1r} + W_{2r})},$$

we find

$$c_{rr} = V(m_r) = \frac{1}{W_{1r} + W_{2r}} - \frac{1}{\Sigma(W_{1r} + W_{2r})} + T_r^2,$$

$$c_{rs} = \operatorname{cov}(m_r, m_s) = -\frac{1}{\Sigma(W_{1r} + W_{2r})} + T_r T_s.$$

This gives

$$V(m_r - m_s) = \frac{1}{W_{1r} + W_{2r}} + \frac{1}{W_{1s} + W_{2s}} + (T_r - T_s)^2,$$

which, with a certain amount of algebra, can be shown to be equal to

$$1 \bigg/ \left[\frac{W_{1r} W_{1s}}{W_{1r} + W_{1s}} + \frac{W_{2r} W_{2s}}{W_{2r} + W_{2s}} \right].$$

This provides a check, since it corresponds to the expression for $V(t_1 - t_2)$ for two experiments, with interchange of treatments and experiments.

Apart from the T terms the variances and covariances of the m's are those of quantities $M_r - \bar{M}$, where the M_r are independently distributed with variances $1/(W_{1r} + W_{2r})$ and $\bar{M}$ is the weighted mean of the M_r. If the true values of the m's are all zero, therefore, the weighted sum of squares

$$\Sigma(W_{1r} + W_{2r}) m_r^2$$

will be distributed as χ^2 with $k - 1$ d.f., apart from the disturbance due to the T terms. The weighted sum of squares has the expectation

$$\Sigma(W_{1r} + W_{2r}) V(m_r) = k - 1 + U\Sigma T_r^2(W_{1r} + W_{2r}).$$

If the T terms are not sufficiently small to be neglected entirely the value of the second term may be deducted before making the χ^2 test, but this correction will usually be small. Since the second term is always positive it need not in any case be calculated if significance is not attained before taking it into account.

Residual χ^2

The general formula for the residual χ^2 is the sum over all cells of

$$\frac{(\text{observed no.} - \text{expected no.})^2}{\text{expected no.}}.$$

If, as in the present case, the data consist of a set of observed proportions, this reduces to

$$S\left\{\frac{(a-m)^2}{m} + \frac{(a-m)^2}{n-m}\right\} = Sn\frac{(p-P)^2}{PQ}.$$

when $a/n = p$ is the observed proportion, $m/n = P = 1-Q$ the expected proportion given by the maximum-likelihood solution, and S denotes summation over all proportions.

We may denote the provisional and working values and weights used for the sth estimates by Y_s, y_s, and W_s. The provisional values Y_s will equal the expected values derived from the $(s-1)$th estimates. We have

$$y_s = Y_s + (p - P_s)\left(\frac{dY}{dP}\right)_s.$$

Consequently the residual χ^2 based on the $(s-1)$th estimates is given by

$$(\chi_r^2)_{s-1} = Sn\frac{(y_s - Y_s)^2}{P_s Q_s}\left(\frac{dP}{dY}\right)_s^2$$

$$= SW_s(y_s - Y_s)^2. \tag{10}$$

The residual χ^2 obtained from the expected values derived from the sth estimates is given by

$$(\chi_r^2)_s = S\{W_{s+1}(y_{s+1} - Y_{s+1})^2\}. \tag{11}$$

If the estimation process is stopped after the sth estimates it will be necessary to calculate specially Y_{s+1}, y_{s+1} and W_{s+1} to evaluate $(\chi_r^2)_s$. Usually, however, $(\chi_r^2)_{s-1}$ will be sufficiently accurate.

Finney (1952) recommends the use of the ordinary formula for the sum of the squares of the residuals when regression coefficients or constants b_1, b_2, ... are fitted by least squares. This is

$$SW(y-Y)^2 = SWy^2 - b_1 B_1 - b_2 B_2 - \dots,$$

where B_1, B_2, ... are the numerical terms of the normal equations with leading terms b_1, b_2, ... A little consideration will show that if the sth estimates of the parameters and the numerical terms of the sth set of normal equations, together with y_s and W_s, are used to calculate the right-hand side of the above equation, we shall obtain exactly

$$SW_s(y_s - Y_{s+1})^2.$$

This, however, is not likely to give a substantially closer approximation than expression (10). The latter is easier to compute, and also has the advantage of revealing aberrant values.

When some of the expectations are very small it is well known that the χ^2 test is un-reliable. In particular, positive deviations from very small expectations will make exces-sively large contributions to the residual χ^2. Under such circumstances Fisher (1950) has shown that $2S\{a \log_e (a/m)\}$, tested against the ordinary χ^2 distribution with the customary number of degrees of freedom, gives a much better measure of the significance of the deviations, and is well fitted to take the place of χ^2 when the expectations are small. An

example of its use is given in example (c) below. This criterion requires somewhat more computation for its evaluation than χ_r^2. since it cannot be simply derived from the transformed values. In our notation the contribution of each observed proportion is

$$2n\{p(\log_e p - \log_e P) + q(\log_e q - \log_e Q)\}.$$

ESTIMATION OF ERROR AND TESTS OF SIGNIFICANCE IN THE *DROSOPHILA* DATA

The above formulae can now be applied to the *Drosophila* data. Taking the values of the weights used for deriving the second estimates (Table 4) we have

$$V(t_1 - t_2) = 1/149 \cdot 96 = 0 \cdot 0817^2.$$

Hence $t_1 - t_2 = +0 \cdot 2937 \pm 0 \cdot 0817$. Reference to the normal probability integral gives $P = 0 \cdot 000163$. Hence the significance of the difference is clearly established. If the actual mutation rates are of interest the standard errors of t_1, t_2 and $\frac{1}{2}(t_1 + t_2)$ will also be required. These are given by

$$V(t_1) = \frac{1}{629 \cdot 6} + \frac{295 \cdot 8^2}{149 \cdot 96 \times 629 \cdot 6^2} = 0 \cdot 0553^2,$$

$$V(t_2) = \frac{1}{629 \cdot 6} + \frac{333 \cdot 8^2}{149 \cdot 96 \times 629 \cdot 6^2} = 0 \cdot 0588^2,$$

$$V\tfrac{1}{2}(t_1 + t_2) = \frac{1}{629 \cdot 6} + \frac{(333 \cdot 8 - 295 \cdot 8)^2}{4 \times 149 \cdot 96 \times 629 \cdot 6^2} = 0 \cdot 0399^2.$$

The test for variation in sensitivity between the experiments is equivalent to testing that the true values of m_r are zero. We have

$$\Sigma(W_{1r} + W_{2r}) m_r^2 = 4 \cdot 745,$$

which has to be compared with χ^2 with 5 d.f. There is therefore no evidence of heterogeneity between experiments. Taking the T_r' into account increases the expectation of the above expression by $0 \cdot 0458$. Clearly the T_r are of no consequence here.

Using expression (10) and the provisional and working values of the second approximation gives a value of χ_r^2 of $5 \cdot 15$. Since the maximum-likelihood solution is very close to the minimum χ^2 solution, it is to be expected that this value of the residual χ^2 will be slightly too large, and in fact the provisional and working values of the third approximation (equivalent to the second approximation estimates) gives a value of $5 \cdot 11$. The value based on the provisional and working values of the second approximation is clearly adequate for practical purposes.

OTHER EXAMPLES

As further examples, which bring out some additional points. we may take the three sets of data given by E. S. Pearson (1950).

(a) Road research data

Table 5 gives the number of cycles and motor-cycles involved in accidents causing personal injury on sections (6 to 15 miles long) of five main roads near London. together with estimates of the miles travelled by these two types of vehicle on these sections. A comparison of the probabilities of being involved in an accident per vehicle mile is required.

Since a vehicle which is involved in an accident is either removed from the road or is again exposed to risk of further accident the number of accidents will follow a Poisson distribution. If the probability of accident per vehicle mile is θ the expected number of accidents per n vehicle miles will be $n\theta$. Following the procedure of the previous example,

we may denote the probability of accident for cycles and motor-cycles on road r by $\mu_r \theta_1$ and $\mu_r \theta_2$. This will make allowances for differences between the different roads which affect both probabilities proportionally.

Table 5. *Road Research data*

Road	Cycles			Motor-cycles		
	Accidents	Miles $\cdot 10^3$	p	Accidents	Miles $\cdot 10^3$	p
1	5	1690	0·0030	4	800	0·0050
2	2	1800	0·0011	2	500	0·0040
3	3	1820	0·0016	1	330	0·0030
4	4	1790	0·0022	3	570	0·0053
5	3	1560	0·0019	5	540	0·0093
	17	8660	0·00196	15	2740	0·00547

The Poisson distribution with mean m can be regarded as the limit of the binomial distribution $(q+p)^n$. in which $p \to 0$ and $n \to \infty$ in such a manner that $np = m$. Consequently, provided the unit of distance travelled is taken sufficiently small for the chance of two accidents occurring in the same unit to be negligible. the data can be treated as if they were binomial. Here a unit of 1000 vehicle miles is sufficiently small. As before the complementary log log transformation is appropriate. Since

$$y = \log_e(-\log_e q) \to \log_e p$$

as $p \to 0$, a table of natural logarithms can be used for the actual transformation. Working values and weighting coefficients can be obtained from Finney's Table IX, which has adequate range for this purpose. Alternatively, if x is the number of occurrences a table of natural logarithms may be used with

$$W = nw = ne^Y,$$
$$y = Y - 1 + x/W.$$

With this procedure there is no need to choose units such that $p = x/n$ is small.

The computations follow the same lines as those of the previous example, and need not be reproduced in detail here. For cycles on the first road, for instance,

$$n = 1690, \quad n' = 5, \quad p = 0\cdot0030, \quad \log_e p = -5\cdot81,$$

and we therefore have:

	Y	y	W
1st estimate	—	−5·81	5
2nd estimate	−6·0	−5·79	4·2

the provisional value $Y = -6\cdot0$ being given by the first estimate of $t_1 + m_1$. For the first estimate $W = n'$, since $w = P = p$.

The values of the first and second estimates of the parameters are shown in Table 6. The second set of estimates differs little from the first and is clearly sufficiently accurate.

The average difference is clearly significant, motor-cycles being estimated as about 2·6 times as liable as cycles to be involved in an accident per 1000 vehicle miles. Taking $\pm 1·65$ times the standard error, we obtain approximate $5\,\%$ lower and upper limits of error of this ratio as 1·5 and 4·7.

The values of the various χ^2 are shown in Table 7. There is no evidence of variation in risk between the different roads, or of any variation of the difference between cycles and motor-cycles.

Table 6. *Road Research data: 1st and 2nd estimates*

	1st	2nd		1st	2nd
t_1	$-6·1597$	$-6·1686$	m_1	$+0·1420$	$+0·1304$
t_2	$-5·1824$	$-5·1973$	m_2	$-0·4930$	$-0·4938$
t_1-t_2	$-0·9773$	$-0·9713$	m_3	$-0·3671$	$-0·3633$
$\Sigma \dfrac{W_1 W_2}{W_1+W_2}$	7·56	7·89	m_4	$-0·0020$	$+0·0117$
S.E. (t_1-t_2)	—	$\pm 0·356$	m_5	$+0·2726$	$+0·2423$

Table 7. *Road Research data: values of χ^2*

	D.F.	χ^2
Average difference (t_1-t_2)	1	7·41
Variation between roads (m's)	4	2·12
Variation in difference (residual)	4	1·40*
Total	9	10·93

* From $SW_2(Y_2-y_2)$. $SW_2(Y_2-y_2)$ gives 1·38.

That there is no appreciable variation between roads can in fact be seen by inspection of the values of p_1 and p_2 in the original data; such variation would result in a positive correlation between the values of p_1 and p_2. This fact suggests a short cut to the whole analysis, since if there is no variation between roads or in the difference between cycles and motor-cycles the data for each type of vehicle may be pooled. A combined test for absence of such variation may be quickly made from the χ^2 for the variation of the ratios for each vehicle separately. These χ^2 can be calculated in the ordinary manner and are found to be:

	D.F.	χ^2
Cycles	4	1·67
Motor-cycles	4	2·01
	8	3·68

Using the pooled data we may estimate immediately that cycles and motor cycles have risks of 0·00196 and 0·00548 per 1000 vehicle miles of being involved in an accident. These agree closely with the estimates of 0·00209 and 0·00553 given by the previous analysis.

The possibility of pooling fragmentary data should always be examined. If pooling is justified the numerical work is considerably reduced and the presentation of the data is simplified. In addition, the amount of information is increased when heterogeneity between the different sets of data can be ignored.

(b) Artificial insemination data

These data (Table 8) are the result of an investigation (Rothschild, 1949) to test whether the passage of a small electric current for the measurement of the activity of the spermatozoa had any adverse effect on the fertility of bull semen. Seven samples were each divided into two parts, one being subjected to the electric current. The half samples were then diluted and used to inseminate heifers. The columns headed 'returns' give the numbers of heifers which did not become pregnant, as judged by their being returned by their owners for reinsemination.

Table 8. Artificial insemination data

Sample	Untreated semen			Treated semen		
	Returns	Total	p	Returns	Total	p
1	5	11	0·455	1	9	0·111
2	3	9	0·333	4	9	0·444
3	3	12	0·25	3	5	0·6
4	6	11	0·545	3	9	0·333
5	4	10	0·4	0	5	0
6	3	11	0·273	1	5	0·2
7	1	4	0·25	3	6	0·5
	25	68	0·368	15	48	0·312

Inspection of the individual p values gives no indication of any heterogeneity between samples. The uniformity of the data is confirmed by the χ^2 between samples for treated and untreated semen separately. The values are:

	D.F.	χ^2
Untreated	6	3·32
Treated	6	7·92
Total	12	11·24

The residual χ^2 is, however, unreliable in this set of data owing to the small expectations of the non-zero classes of explosive A (ignitions) in Exps. 4 and 5. The likelihood criterion $-2N\{a \log_e (a/m)\}$ has the value of 13·41, giving a significance level $P = 0·020$. This may be compared with the significance level obtained by combining the data into the two groups, namely, Exps. 4 and 5 and Exps. 1, 2, 3 and 6. (This particular combination is valid for the purpose of a test of significance, since 4 and 5 are the two experiments with low values of n'/n.) A $2 \times 2 \times 2$ table then results. For the interaction of such a table, as Bartlett (1935) has shown, the exact probabilities of the various possible occurrences with all marginal totals fixed can be calculated in a similar manner to that followed in a 2×2 table. The resultant test is in fact a test for the constancy of the differences of the logits.* Bartlett's test gives a probability of 0·0147 of two or more of the three ignitions in Exps. 4 and 5 being associated with explosive A. Bearing in mind (a) that practically the whole of the discrepancy is in this test concentrated in a single degree of freedom, but that, on the other hand (b) there is a large effect of discontinuity, not present in the likelihood criterion, the agreement between the two tests is satisfactory.

The indication of the residual χ_r^2 that these experiments are anomalous is therefore confirmed, though the significance level is much less than that given by the χ^2 value. This result is not, however, of great consequence, since it merely provides evidence of the incorrectness of the supposition that the difference in logits between the two explosives is constant over the whole of the probability scale. Our conclusions from these data should be that explosive B produces significantly fewer ignitions than explosive A when conditions are such that the probability of ignition is substantial, but that there is no evidence of any difference when the probability is low. Substantially more data at low probabilities would be required, however, to reach any firm conclusions as to what is really happening in this region.

SUMMARY

The analysis of sets of quantal experiments with two treatments by the use of transformations and maximum likelihood is discussed, and illustrated by examples, one of which is also treated by large-sample methods. Methods are given for (a) obtaining an estimate of the treatment difference in terms of the transformed variate, together with its standard error, (b) testing for differences in sensitivity between the different experiments, and (c) testing for variation in the treatment difference from experiment to experiment. The appropriateness of various transformations under different circumstances is also considered.

REFERENCES

BARTLETT, M. S. (1935). *J. R. Statist. Soc. Suppl.* **2**, 248.
BONNIER, G. & LÜNING, K. G. (1953). *Hereditas, Lund.*, **39**, 193.
COCHRAN, W. G. (1942). *Iowa State Coll. J. Sci.* **16**, 421.
COCHRAN, W. G. (1952). *Ann. Math. Statist.* **23**, 315.
COCHRAN, W. G. (1954). *Biometrics*, **10**, 417.
DYKE, G. V. & PATTERSON, H. D. (1952). *Biometrics*, **8**, 1.
FINNEY, D. J. (1952). *Statistical Method in Biological Assay.* London: Griffin.
FISHER, R. A. (1932). *Statistical Methods for Research Workers* (4th ed.). Edinburgh: Oliver and Boyd.

* Simpson (1951), following a suggestion by Bartlett, has drawn attention to certain errors in the latter's paper, which do not, however, affect the test for interaction. Lancaster (1951) has criticized Bartlett's test on what seems to me the invalid ground that the components of χ^2 associated with this test are not additive.

FISHER, R. A. (1950). *Biometrics*, **6**, 17.

FISHER, R. A. & YATES, F. (1938). *Statistical Tables for Biological, Agricultural and Medical Research* (4th ed. 1953). Edinburgh: Oliver and Boyd.

JOLLY, G. M. (1950). *Ann. Appl. Biol.* **37**, 597.

LANCASTER, H. O. (1949). *Biometrika*, **36**, 370.

LANCASTER, H. O. (1951). *J. R. Statist. Soc.* B, **13**, 242.

PEARSON, E. S. (1950). *Biometrika*, **37**, 383.

PEARSON, K. (1933). *Biometrika*, **25**, 379.

ROTHSCHILD, LORD (1949). *J. Agric. Sci.* **39**, 294.

SIMPSON, E. H. (1951). *J. R. Statist. Soc.* B, **13**, 238.

YATES, F. (1953). *Sampling Methods for Censuses and Surveys* (2nd ed.). London: Griffin.

YATES, F. (1955). *Biometrika*, **42**, 404.

YATES, F. & HALE, R. W. (1939). *J. R. Statist. Soc. Suppl.* **6**, 67.

A NOTE ON THE APPLICATION OF THE COMBINATION OF PROBABILITIES TEST TO A SET OF 2×2 TABLES

By F. YATES, F.R.S.

Rothamsted Experimental Station

INTRODUCTION

The combination of probabilities test, or probability integral test, as it is also called, has been commonly used to give an overall test of significance for a set of 2×2 contingency tables for which pooling is not regarded as admissible because of heterogeneity of the data. Data of this type can, in my opinion, be more effectively analysed by the method of maximum likelihood, using some appropriate transformation; the methods of analysis are described in a separate paper (Yates, 1955). Cases will, however, arise where a quick, possibly preliminary, test of significance is required, and for this purpose the combination of probabilities test or some analogous test may be regarded as adequate. It therefore seems worth putting on record certain points concerning tests of this type.

The combination of probabilities test requires the calculation of a significance level P for each of the k tables separately. The test criterion $S(- 2 \log_e P)$ is then compared with the χ^2 distribution with $2k$ degrees of freedom. We usually require to test whether the data collectively show evidence of deviation in a common direction, and in such cases the significance level P must be that of the appropriate single tail of the corresponding distribution.

The test is exact if the basic P distributions are continuous with uniform probability, i.e. if each P can be regarded as uniformly distributed between 0 and 1 when the hypothesis to be tested holds. If the P's are derived from 2×2 contingency tables with small marginal totals, however, the distributions of P are by no means continuous. As has been pointed out by Pearson (1950) they can be made continuous with uniform probability by the following device. If for any particular 2×2 table the probability derived from the hypergeometric series is $P_a(H)$ that the number in a particular cell is $\geqslant a$, and the probability of this number being $\geqslant (a+1)$ is $P_{a+1}(H)$, then if a is the number observed we may use a value $P_{a+\theta}(H)$ instead of $P_a(H)$, where $P_{a+\theta}(H)$ is a random selection with uniform probability from values in the range $P_a(H)$ to $P_{a+1}(H)$.

Since all the values of P will be decreased by applying this adjustment, the value of $S(- 2 \log_e P)$ will be increased. Without the adjustment, therefore, we shall consistently over-estimate the combined probability of obtaining a set of P's whose product is as small as or smaller than that observed. This confirms the result obtained by Lancaster (1949) in the course of a detailed investigation on the combination of probabilities from discrete distributions.

To overcome this difficulty Lancaster suggested that the average value of $- 2 \log_e P_{a+\theta}(H)$ be used instead of the value $- 2 \log_e P_a(H)$. This average value he denoted by χ_m^2, but to avoid confusion with the various χ^2 with 1 d.f. appertaining to the individual experiments, we will here denote it by $- 2 \log_e P_{a+}(H)$. We have

$$- 2 \log_e P_{a+}(H) = \chi_m^2 = E\{ - 2 \log_e P_{a+\theta}(H)\}$$
$$= 2 - 2\{P_a(H) \log_e P_a(H) - P_{a+1}(H) \log_e P_{a+1}(H)\} \{P_a(H) - P_{a+1}(H)\}.$$

As a further alternative, to save computation, Lancaster suggested what he called the median value χ_m^2, which he defined by

$$\chi_m'^2 = -2\log_e \tfrac{1}{2}\{P_a(H) + P_{a+1}(H)\} \quad P_{a+1}(H) \neq 0,$$
$$= 2 - 2\log_e P_a(H), \qquad\qquad P_{a+1}(H) = 0.$$

Neither of Lancaster's procedures, of course, gives an exact χ^2 distribution with $2k$ d.f., but the departures will be very slight in all cases likely to be met with in practice. By Pearson's procedure an exact χ^2 distribution can be generated, but this introduces an arbitrary element which, as Pearson himself recognized, may be regarded as altogether too high a price to pay for a formally exact test.

USE OF χ^2 TO DETERMINE THE VALUES OF P

If the number of terms in the hypergeometric series appertaining to a single experiment is at all large the calculation of the exact probabilities is a tedious matter. An alternative procedure is to calculate the values of χ^2 for the individual experiments without the correction for continuity, and use these to determine the corresponding one-tail probabilities. Cochran (1942, 1952), correcting a misleading statement by myself (1934), pointed out that the continuity correction should not be applied when summing χ^2 from a number of 2×2 tables, though the reason he gives in his 1952 paper, that 'the correction has a tendency to over-correct' and that 'the over-correction mounts up in a disconcerting manner' is not, as we have seen above, the basic one. Lancaster discussed the use of χ^2 without correction for continuity for the combination of probabilities test, but confined his attention to the calculation of the two-tail probabilities. Unfortunately, as already mentioned, this is not what is usually required in practice, and Lancaster's argument is not applicable to the one-tail case.

The question of the accuracy of the χ^2 approximation, therefore, still requires investigation. When χ^2 corrected for continuity, χ_c^2 say, gives a good approximation to the one-tail probabilities, it may be expected that the use of χ^2 without correction for continuity, χ_u^2 say, for the calculation of the probabilities for combination will give a reasonable approximation to the test based on the exact probabilities and $-2\log_e P_{a\cdot}(H)$ as defined above. For if $P_a(\chi_u)$ is the one-tail probability given by χ_u^2, when the value a is observed, and $P_a(\chi_c)$ is the corresponding probability given by χ_c^2, then $P_a(\chi_u)$ is intermediate between $P_a(\chi_c)$ and $P_{a+1}(\chi_c)$, and will not differ greatly from $P_{a\cdot}(\chi_c)$.

It is well known, however, that if the relevant hypergeometric distribution is very asymmetrical $P_a(\chi_c)$ is not a good approximation to $P_a(H)$ (Yates, 1934). Consequently, if the data are such that the hypergeometric distributions for the separate tables are very asymmetrical, and particularly if the asymmetry is in the same sense for all tables, the use of χ^2 without correction for continuity instead of the exact probabilities may be expected to give misleading results.

Some comparisons made for actual 2×2 tables, however, are reassuring. Tables having the following marginal totals (treatments 1 and 2) were taken:

Table A

	1	2	
+	a		8
−			192
	100	100	200

Table B

	1	2	
+	a		8
−			142
	50	100	150

Table C

	1	2	
+	a		16
−			224
	40	200	240

Table A gives a symmetrical distribution and Tables B and C give increasing degrees of asymmetry. (No particular merit is claimed for these tables; A was originally chosen for another purpose, and B and C were taken to give distributions differing from A mainly in degree of symmetry.)

For these three tables the values of $-2 \log_e P_{a+}(H)$ and $-2 \log_e P_a(\chi_u)$ were calculated for all possible values of a. For Tables B and C there are two sets of values depending on the direction in which the deviations are measured. The results are shown in Table 1. In distribution A all the individual values except that for $a = 8$ (which will only occur in 1 out

Table 1. *Comparison of* $-2 \log_e P_{a+}(H)$ *and* $-2 \log_e P_a(\chi_u)$

Distribution A

a	$p(a)$	Test for 1 > 2	
		$-2 \log_e P_{a+}(H)$	$-2 \log_e P_a(\chi_u)$
0	0·003377	0·004	0·004
1	0·029052	0·036	0·030
2	0·107092	0·181	0·155
3	0·220948	0·583	0·536
4	0·279062	1·413	1·386
5	0·220948	2·842	2·894
6	0·107092	5·055	5·195
7	0·029052	8·332	8·382
8	0·003377	13·383	12·484
	1·000000	2·000	2·004

Distribution B

a	$p(a)$	Test for 1 > 2		Test for 1 < 2	
		$-2 \log_e P_{a+}(H)$	$-2 \log_e P_a(\chi_u)$	$-2 \log_e P_{(8-a)+}(H)$	$-2 \log_e P_{(8-a)}(\chi_u)$
0	0·035397	0·036	0·040	8·682	7·833
1	0·152244	0·239	0·210	4·571	4·616
2	0·277764	0·805	0·724	2·302	2·384
3	0·280688	1·905	1·839	1·021	1·017
4	0·171775	3·662	3·767	0·371	0·330
5	0·065168	6·178	6·646	0·102	0·073
6	0·014962	9·581	10·56	0·019	0·010
7	0·001900	14·11	15·56	0·002	0·001
8	0·000102	20·38	21·67	0·000	0·000
	1·000000	2·000	2·021	2·000	1·989

Table 1. (cont.)

Distribution C

| a | p(a) | Test for 1 > 2 | | Test for 1 < 2 | |
		$-2 \log_e P_{a+}(H)$	$-2 \log_e P_a(\chi_u)$	$-2 \log_e P_{(a-)}(H)$	$-2 \log_e P_{(a-)}(\chi_u)$
0	0·048696	0·049	0·065	8·044	6·881
1	0·168463	0·288	0·264	4·190	4·182
2	0·264922	0·874	0·776	2·152	2·268
3	0·251227	1·907	1·791	1·011	1·050
4	0·160692	3·426	3·460	0·416	0·340
5	0·073459	5·442	5·890	0·144	0·108
6	0·024809	7·955	9·148	0·041	0·021
7	0·006309	10·96	13·27	0·009	0·002
8	0·001220	14·46	18·29	0·002	0·000
9	0·000180	18·47	24·23	...	...
10	0·000020	...	...	...	...
11	0·000002	...	...	...	...
...	...	...	...	...	...
	0·999999	2·000	2·030	2·000	1·975

of 300 experiments) agree to within 0·15, and the expectation of $-2 \log_e P_a(\chi_u)$ is almost exactly 2. Distributions B and C show considerably greater individual differences, as is to be expected, but the large differences are again confined to the tails, where they are of no great consequence; for if values occur in the tail in any substantial proportion of the experiments the significance of the combined results will not in any case be in doubt. The expectations deviate in opposite directions for the two types of test but are still very close to 2.

On this last point it may be noted that Lancaster defines (without comment) what he calls the crude χ^2 of a 2×2 table as

$$\chi^2 = \frac{(N-1)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}.$$

The customary form, which has been used in this paper for χ_u^2, has the factor N instead of $N-1$ in the numerator. Lancaster's form has the property, which is useful for the combination of two-tail tests, that the expectation is 1. This, of course, does not make the expectation of $-2 \log_e P_a(\chi)$ equal to 2, but it does in fact bring it closer to 2 in the case of symmetrical distributions. It can do little, however, to improve the expectations for asymmetrical distributions, since the deviations are of opposite sign for the two types of test.

From the above comparisons we may conclude that the determination of the one-tail probabilities from χ^2 without correction for continuity will be satisfactory in cases likely to be met with in practice, even when the expectations in the individual experiments are quite small.

Biom. 42

27

VARIANTS OF THE TEST

The use of values of χ^2 for 2 d.f. for the combination of probabilities is to a certain extent arbitrary. It has the convenience that the values are easily calculated, and the use of a function of the product of the probabilities has a certain intuitive appeal, but the method would work equally well with other basic numbers of degrees of freedom. If, for instance, the values of χ^2 for 1 d.f. corresponding to the P's are summed then in the absence of association the sum will be distributed as χ^2 for k d.f. Equally—and in the type of data we are considering this procedure requires even less computation—the values of χ_u in the absence of association approximate to normal deviates with zero mean and unit standard deviation, and their sum is therefore a normal deviate with a standard deviation of $\sqrt{k}$ (see, for example, Cochran, 1954).

The calculations for these variants for the *Drosophila* data analysed in the paper referred to (Yates, 1955) are exhibited in Table 2. The combination of probabilities test (column 4) gives a significance level (one tail) of $P = 0.000610$, whereas the use of χ^2 for 1 d.f. instead of 2 d.f. (column 5) gives $P = 0.000760$, and $S(\chi_u)$ (column 2) gives $P = 0.000673$.

Table 2. *Calculation of significance levels for the* Drosophila *data*

Exp.	χ_u^2	χ_u	$P_a(\chi_u)$	$-2 \log_e P_a(\chi_u)$	$\chi_1^2\{P_a(\chi_u)\}$
	(1)	(2)	(3)	(4)	(5)
1	0·4448	0·667	0·2524	2·753	1·309
2	4·2274	2·056	0·01989	7·835	5·420
3	8·0065	2·830	0·002327	12·126	9·272
4	0·0418	−0·204	0·5808	1·087	0·305
5	0·0938	0·306	0·3798	1·936	0·771
6	4·8325	2·198	0·01397	8·542	6·042
	17·6468	7·853	—	34·279	23·119
P	0·00718	0·000673	—	0·000610	0·000760

These values are in close agreement. Indeed, except when the data from the different experiments are mutually contradictory no great differences are to be expected between the different possible methods that suggest themselves, since they merely demarcate equi-probable contour surfaces in the multiple P space which are not markedly different. Nevertheless, alternative tests which are merely variants based on the same information are likely to give rise to confusion, and it is therefore recommended that some standard procedure is adopted whenever a test based on the levels of significance in the individual experiments or sets of observations is required. If the information presented to the statistician consists of the relevant probabilities the combination of probabilities test commends itself both on historical grounds and because of its simplicity and intuitive appeal. If a one-tail test of a set of 2×2 tables based on the χ^2 values for each table is required, however, there seems little point in transforming the values of χ_u into probabilities and then re-transforming these into χ^2 values. It is therefore recommended that in such cases $S(\chi_u)$ be adopted as the standard test criterion.

In contrast with the above values of P it may be noted that $S(\chi_a^2)$ (column 1), which in the absence of association would also be approximately distributed as χ^2 with 6 d.f., gives a value of $P = 0.00718$. This considerably larger value of P is to be expected, since the direction of the deviations is not taken into account.

EFFICIENCY OF THE TEST

The significance levels given by the combination of probabilities test and by the ratio of $t_1 - t_2$ to its standard error in the maximum-likelihood solution for the four examples of the other paper (Yates, 1955) are as follows:

	Combination of probabilities	Maximum likelihood
Drosophila data	0.000610	0.000163
Road research data	0.00691 *	0.00317
Artificial insemination data	0.64 *	0.79
Safety in mines data	0.000517 *	0.000402

* Computed from the values of $S(\chi_m^2)$ given by Pearson (1950).

In all three cases in which significance is attained the maximum-likelihood solution gives a higher level of significance than the combination of probabilities test. In two cases the difference is substantial. This suggests that the combination of probabilities test is not very efficient, though the differences may in part be due to chance causes or to inadmissible approximations in one or both the tests.

Lancaster compared the power of the combination of probabilities test, the direct summation of χ^2 test, and the test derived from the pooled data by numerical methods, taking the case of the binomial distribution $(q + p)^5$ with $p = 0.5$ as the null hypothesis and enumerating all possible events. He made a similar comparison of the first two tests and some variants on a 2×2 table. Unfortunately, however, he used two-sided tests. This considerably reduces the efficiency of the combination of probabilities or summation of χ^2 tests, and his estimates of the relative power of the three tests are therefore not really relevant to the situations met with in practice.

A full investigation of the efficiency of the combination of probabilities test is beyond the scope of this note, but it is to be expected on general grounds that the combination of probabilities test will be somewhat inefficient. In the first place no direct cognizance is taken of the number of observations and other factors that affect the amount of information given by the different experiments. It is true that the more accurate experiments will, on the average, yield results of greater significance when there is a real difference. But if, for example, of two experiments A and B, A yields 10 times the amount of information given by B, we should naturally be inclined to give more weight to a significant result from experiment A than to one from experiment B. But on the combination of probabilities test the result $P_A = 0.025$, $P_B = 0.3$, for instance, will give exactly the same combined level of significance (5% single tail), as $P_A = 0.3$, $P_B = 0.025$.

In the second place even if all the experiments yield the same amount of information this information is not in general combined efficiently by the combination of probabilities test.

27-2

As a simple example we may consider the case of a set of quantitative experiments which furnish normally distributed estimates $x_1, x_2, \ldots, x_k$ of a constant treatment difference μ, all having the same (known) variance σ^2. The efficient (indeed sufficient) combined estimate of μ will then be $\bar{x}$, and this will be normally distributed with known variance $\sigma^2 k$ and can therefore be used to give an exact overall P, P_L say. The exact level of significance can also be calculated for each experiment separately, and these levels can then be combined to give an overall P, P_C say. It is easily seen that in particular cases P_C may differ considerably from P_L.

Table 3. *Comparison of significance levels given by maximum likelihood (P_L) and combination of probabilities (P_C) in a sampling experiment*

$P(\bar{x}-0.9)$	P_L	P_C	$\log_{10}(P_C/P_L)$
0.9987	0.839	0.820	−0.010
0.729	0.0808	0.0955	0.073
0.719	0.0764	0.0460	−0.220
0.702	0.0694	0.0908	0.117
0.663	0.0559	0.0409	−0.136
0.659	0.0548	0.0296	−0.267
0.641	0.0495	0.0512	0.015
0.618	0.0436	0.0838	0.284
0.540	0.0281	0.0145	−0.287
0.480	0.0197	0.0367	0.270
0.405	0.0122	0.0318	0.416
0.405	0.0122	0.0132	0.034
0.334	0.00734	0.0216	0.469
0.258	0.00391	0.0100	0.408
0.233	0.00307	0.00394	0.108
0.142	0.00104	0.00118	0.055
0.131	0.000874	0.00185	0.326
0.111	0.000619	0.00287	0.666
0.090	0.000404	0.00132	0.514
0.038	0.0000784	0.0000973	0.094
			0.146

A test of the relative performance of the two tests was made for a specific example of this type with $k = 5$, $\sigma = 1$, and a true treatment difference of 0.9. With these values $\bar{x}$ will attain significance at the 2.2 % level (single tail) in 50 % of all sets of five experiments. The actual levels attained by $\bar{x}$ and by the combination of probabilities test were calculated for twenty such sets, using random normal deviates nos. 6551–6650, each increased by 0.9, given by Wold (1948). The results are shown in Table 3. $x - 0.9$ will be normally distributed about zero mean with standard deviation $1/\sqrt{5}$. The probability $P(x - 0.9)$ of getting a value of $\bar{x} - 0.9$ greater than the observed value was calculated for each set, and the results have been arranged in order of magnitude of this probability. With a large number of sets there would be a uniform distribution of $P(x - 0.9)$ over the range 0–1.

The results show clearly that P_L tends to be less than P_C, and that the difference is greater in the experiments with small $P(x - 0.9)$ which, of course, also have small P_L. The average

value of $\log_{10} P_c P_L$ is 0·146. In other words, the significance levels are on the average in the ratio of 1·40 : 1. There is thus a clear loss of efficiency with P_c. Equally important, there are considerable discrepancies in individual sets of experiments, leading to a conflict of evidence from the same body of data which arises solely from the use of an inefficient test.

SUMMARY

It is shown that χ^2 without correction for continuity will give one-tail probabilities for 2×2 tables which may be safely combined in most cases likely to be met with in practice. The summation of the corresponding signed values of χ gives a rapid method of combination. Reasons are given for believing that combination of probabilities tests are not likely to be very efficient, and this conclusion is demonstrated by a small sampling experiment.

REFERENCES

COCHRAN, W. G. (1942). *Iowa State Coll. J. Sci.* **16**, 421.
COCHRAN, W. G. (1952). *Ann. Math. Statist.* **23**, 315.
COCHRAN, W. G. (1954). *Biometrics*, **10**, 417.
LANCASTER, H. O. (1949). *Biometrika*, **36**, 370.
PEARSON, E. S. (1950). *Biometrika*, **37**, 383.
WOLD, H. (1948). *Random Normal Deviates*. Tracts for computers, no. XXV. Cambridge University Press.
YATES, F. (1934). *J. R. Statist. Soc. Suppl.* **1**, 217.
YATES, F. (1955). *Biometrika*, **42**, 382.

A TEST FOR HOMOGENEITY OF THE MARGINAL DISTRIBUTIONS IN A TWO-WAY CLASSIFICATION

By ALAN STUART

Division of Research Techniques, London School of Economics

1. INTRODUCTION

There are several circumstances in which we may wish to test the homogeneity of the two sets of marginal probabilities in a two-way classification. For example, a sample from a bivariate distribution (say height of father, height of son) may be classified into a two-way table with identical (height) groupings in each margin. Or a similar classification may be possible for a non-measurable variable (say strength of right hand − strength of left hand). Again, in surveys of the same sample (a 'panel') on two different occasions, the inter-relation of the results on the two occasions may be displayed in a two-way table, with one margin corresponding to each occasion. In all these cases, the question may arise: are the two sets of marginal probabilities identical?

If the variable is measurable, we may test the difference between the *means* of the two marginal distributions by a large-sample standard-error test. However, we may be interested in the overall distributions, rather than only in their means. For the more stringent hypothesis of homogeneity, a test exists if we have two completely independent samples, when an ordinary χ^2 test of homogeneity may be applied (Cramer, 1946, p. 445). This test does not meet the essentially bivariate situations described above, where non-independence of the marginal distributions is a fundamental feature of the problem.

When the classification is a double dichotomy, the problem of testing marginal homogeneity is simple, and its solution is a special case of the large-sample solution of the more general 2^K classification problem given by Cochran (1950). Bowker (1948) gave a large-sample test for complete symmetry in a two-way classification, a more restrictive hypothesis which is concerned with the entire set of probabilities in the classification, and not only with the marginal probabilities as we are here. In the present paper, a large-sample test for marginal homogeneity is derived and illustrated.

2. VARIANCES AND COVARIANCES

With the usual notation for an $m \times m$ table, we denote the probability of falling into the ith row and jth column by p_{ij}, and define the marginal probabilities by

$$p_{i.} = \sum_j p_{ij}, \quad p_{.j} = \sum_i p_{ij},$$

while obviously

$$\sum_i \sum_j p_{ij} = \sum_i p_{i.} = \sum_j p_{.j} = 1.$$

The corresponding sample numbers are denoted by n_{ij}, $n_{i.}$, $n_{.j}$, while the total sample size is n.

By standard multinomial theory, we have for the means, variances and co-variances of the n_{ij}

$$E(n_{ij}) = np_{ij},$$
$$V(n_{ij}) = np_{ij}(1-p_{ij}),$$
$$C(n_{ik}, n_{lj}) = -np_{ik}p_{lj} \quad (i \neq l \text{ and or } k \neq j). \tag{1}$$

and also

$$E(n_{i.}) = np_{i.},$$
$$V(n_{i.}) = np_{i.}(1-p_{i.}),$$
$$C(n_{i.}, n_{j.}) = -np_{i.}p_{j.} \quad (i \neq j),$$
$$C(n_{.i}, n_{.j}) = -np_{.i}p_{.j} \quad (i \neq j). \tag{2}$$

We now require

$$C(n_{i.}, n_{.j}) = V(n_{ij}) + \sum_{k,l} C(n_{ik}, n_{lj}),$$

where $i \neq l$ and/or $k \neq j$. From (1), this is

$$C(n_{i.}, n_{.j}) = n\{p_{ij}(1-p_{ij}) - \sum_{k,l} p_{ik}p_{lj}\},$$
$$= n(p_{ij} - p_{i.}p_{.j}), \tag{3}$$

on taking the term in p_{ij}^2 into the summation. We now define the statistics

$$d_i = n_{i.} - n_{.i} \quad (i = 1, 2, \ldots, m).$$

Since the likelihood-ratio principle yields an intractable result in this case, it seems natural to use the d_i as the basis for a test concerning the differences between the corresponding marginal probabilities. For their means, variances and co-variances, we have, from (2) and (3), the exact results:

$$E(d_i) = E(n_{i.} - n_{.i}) = n(p_{i.} - p_{.i}), \tag{4}$$
$$V(d_i) = V(n_{i.}) + V(n_{.i}) - 2C(n_{i.}, n_{.i})$$
$$= n\{(p_{i.} + p_{.i} - 2p_{ii}) - (p_{i.} - p_{.i})^2\}. \tag{5}$$

Also, for $i \neq j$

$$C(d_i, d_j) = C(n_{i.}, n_{j.}) + C(n_{.i}, n_{.j}) - \{C(n_{i.}, n_{.j}) + C(n_{.i}, n_{j.})\}$$
$$= -n\{(p_{ij} + p_{ji}) + (p_{i.} - p_{.i})(p_{j.} - p_{.j})\}. \tag{6}$$

3. TESTING THE HYPOTHESIS OF HOMOGENEITY

Suppose that we wish to test the hypothesis

$$H_1: \quad (p_{i.} - p_{.i}) = \Delta_i \quad (i = 1 \text{ to } m),$$

where, of course, we must have

$$\sum_{i=1}^{n} \Delta_i = 0.$$

(4), (5) and (6) then become

$$E(d_i \mid H_1) = n\Delta_i,$$
$$V(d_i \mid H_1) = n\{(p_{i.} + p_{.i} - 2p_{ii}) - \Delta_i^2\},$$
$$C(d_i, d_j \mid H_1) = -n\{(p_{ij} + p_{ji}) + \Delta_i \Delta_j\} \quad (i \neq j), \tag{7}$$

and in the special case of particular interest here, where we have

$$H_0: \quad \Delta_i = 0 \quad (\text{all } i),$$

(7) becomes

$$E(d_i \mid H_0) = 0,$$

$$\left.\begin{array}{l} V_{ii} = V(d_i \mid H_0) = n(p_{i.} + p_{.i} - 2p_{ii}), \\ V_{ij} = C(d_i, d_j \mid H_0) = -n(p_{ij} + p_{ji}) \quad (i \neq j). \end{array}\right\} \tag{8}$$

We have deliberately refrained from replacing $p_{.i}$ by $p_{i.}$ in V_{ii}, as we could have done, since if, as will generally be the case, the true probabilities are unknown, the maximum-likelihood estimators of the variances and co-variances in (8) are

$$\left.\begin{array}{l} \hat{V}_{ii} = n_{i.} + n_{.i} - 2n_{ii}, \\ \hat{V}_{ij} = -(n_{ij} + n_{ji}). \end{array}\right\} \tag{9}$$

It is well known that the n_{ij} have a limiting multinormal distribution with moments given by (1). It follows that the $n_{i.}$ and $n_{.j}$ will also be jointly normally distributed, since they are linear functions of the n_{ij}, and finally, by the same argument, that the variates d_i ($i = 1$ to m) will be asymptotically multinormally distributed with moments given by (7) in general, and by (8) on the null hypothesis. To a further degree of approximation, the same result holds if (9) is used as an estimator of (8).

Now it is a standard result that in the exponent, say $-\frac{1}{2}Q$, of a multinormal distribution, Q is distributed like χ^2 with degrees of freedom equal in number to the rank of the distribution. Furthermore, Q in this exponent is simply a quadratic form in the variables. Thus, on H_0,

$$Q = \sum_{i,j=1}^{m} a_{ij} d_i d_j \tag{10}$$

is distributed in the limit like χ^2 with $(m-1)$ degrees of freedom, the rank of the distribution being $(m-1)$ since $\sum_{i=1}^{n} d_i = 0$, and in general this is the only constraint on the d_i. If the distribution were non-singular, the coefficients a_{ij} would simply be the elements of the inverse of the variance-covariance matrix (V_{ij}). As a result of the singularity, (V_{ij}) itself is also singular, and so we cannot invert it.

However, any marginal distribution of a multinormal distribution is itself normal, so that we may eliminate the redundant variate (say the last) and obtain the result that

$$Q = \sum_{i,j=1}^{m-1} V^{ij} d_i d_j \tag{11}$$

is distributed in the limit like χ^2 with $(m-1)$ degrees of freedom, $(m-1)$ still being the rank of the distribution. If we replace the V^{ij} in (11) by their maximum-likelihood estimators given by (9), the same asymptotic result holds. In (11), (V_{ij}) is now understood to be defined for $i, j = 1$, to $m-1$ only, and (V^{ij}) is its inverse.

The fact that (11) takes no explicit account of d_m lends it an appearance of arbitrariness, but although the values of the terms in Q are changed by eliminating some other d_i instead of d_m, their sum Q is uniquely determined. This is in virtue of the fact that since any $(m-1)$ of the d_i uniquely determine the other one, Q must be, apart from constants, the log likelihood of the complete set of d_i, irrespective of which of the d_i is omitted in Q. In point of fact, Q could be expressed as a function of all the m values of d_i and the complete matrix

(V_{ij}), but this merely complicates the computational procedure and makes no difference to the result.

Finally, we must determine the appropriate critical region of the distribution of Q. (7) shows that, when H_0 is not true, the variance-covariance matrix of the d, may be written

$$(V_{ij}) = n(w_{ij}),$$

where (w_{ij}) is independent of n. The inverse matrix is therefore

$$(V^{ij}) = \frac{1}{n}(w^{ij}).$$

Thus the expected value of Q in (11) is

$$E(Q \mid H_1) = \sum_{i,j=1}^{m-1} \frac{1}{n} w^{ij} E(d_i d_j)$$

$$= \sum \frac{1}{n} w^{ij} \{C(d_i, d_j) + E(d_i) E(d_j)\},$$

so that, using (7),
$$E(Q \mid H_1) = O(n). \tag{12}$$

On the other hand, the limiting χ^2 distribution of Q on H_0 gives us

$$E(Q \mid H_0) = m - 1, \tag{13}$$

independent of n. Thus, with increasing n, the difference between (12) and (13) will exceed any bound, and the appropriate critical region for our large-sample test is the upper tail of the distribution of Q. The same conclusion holds if (V_{ij}) is replaced by its maximum-likelihood estimator $(\hat{V}_{ij})$.

4. COMPUTATIONAL PROCEDURE: AN EXAMPLE

We now set out the computation necessary for the test, and give an illustration of its use.

(1) Form the matrix $(\hat{V}_{ij})$ given by (9), for $i.j = 1$ to $(m-1)$.

(2) Invert $(\hat{V}_{ij})$ to obtain $(\hat{V}^{ij})$.

(3) Compute $Q = \sum_{i,j=1}^{m-1} \hat{V}^{ij} d_i d_j$, where $d_i = n_{i.} - n_{.i}$, and test in the χ^2 distribution with $(m-1)$ degrees of freedom, the critical region being the upper tail.

Consider the following example, which uses data quoted by Stuart (1953).

7477 women aged 30–39: unaided distance vision

Left eye Right eye	Highest grade	Second grade	Third grade	Lowest grade	Total
Highest grade	1520	266	124	66	1976
Second grade	234	1512	432	78	2256
Third grade	117	362	1772	205	2456
Lowest grade	36	82	179	492	789
Total	1907	2222	2507	841	7477

We have
$$\cdot d_1 = 1976 - 1907 = 69,$$
$$d_2 = 2256 - 2222 = 34,$$
$$d_3 = 2456 - 2507 = -51,$$

and we check that $d_4 = 789 - 841 = -52$ equals the sum $(d_1 + d_2 + d_3)$, negatively signed. Since d_1 and d_2 are positive, while d_3 and d_4 are negative, one may ask whether the sight of the right eye may really be regarded as better than that of the left eye for the population from which this is a sample. The estimated variance matrix of d_1, d_2, d_3 is obtained by use of (9), as

$$\hat{V}_{11} = (266 + 124 + 66) + (234 + 117 + 36) = 843,$$
$$\hat{V}_{12} = \hat{V}_{21} = -(234 + 266) = -500,$$

and so on, giving
$$(\hat{V}_{ij}) = \begin{pmatrix} 843 & -500 & -241 \\ -500 & 1454 & -794 \\ -241 & -794 & 1419 \end{pmatrix}.$$

The inverse is obtained directly as
$$(\hat{V}^{ij}) = 10^{-6} \begin{pmatrix} 2482 & 1560 & 1295 \\ 1560 & 1972 & 1368 \\ 1295 & 1368 & 1690 \end{pmatrix},$$

and we have, for our χ^2 statistic with 3 degrees of freedom,
$$Q = 10^{-6}\{(2482 \times 69^2) + (1972 \times 34^2) + (1690 \times 51^2) + 2(1560 \times 69 \times 34)$$
$$- 2(1295 \times 69 \times 51) - 2(1368 \times 34 \times 51)\}$$
$$= 11{\cdot}96.$$

The 1 % point for the χ^2 distribution with 3 degrees of freedom is $11{\cdot}345$, so we conclude that the result is significant of a difference in the distribution of sight in right and left eye.

If, instead, of eliminating d_4, we had eliminated, say, d_1 we should have obtained as above:

$$(\hat{V}_{ij}) = \begin{pmatrix} 1454 & -794 & -160 \\ -794 & 1419 & -384 \\ -160 & -384 & 646 \end{pmatrix},$$

$$(\hat{V}^{ij}) = 10^{-6} \begin{pmatrix} 1332 & 995 & 921 \\ 995 & 1583 & 1187 \\ 921 & 1187 & 2482 \end{pmatrix},$$

and finally
$$Q = 10^{-6}\{(1332 \times 34^2) + (1583 \times 51^2) + (2482 \times 52^2) + 2(1187 \times 51 \times 52)$$
$$- 2(995 \times 34 \times 51) - 2(921 \times 34 \times 52)\}$$
$$= 11{\cdot}96$$
as before.

REFERENCES

BOWKER, ALBERT H. (1948). A test for symmetry in contingency tables. *J. Amer. Statist. Ass.* **43**, 572–4.

COCHRAN, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, **37**, 256–66.

CRAMER, HARALD (1946). *Mathematical Methods of Statistics*. Princeton University Press.

STUART, A. (1953). The estimation and comparison of strengths of association in contingency tables. *Biometrika*, **40**, 105–10.

DISTRIBUTIONS OF KENDALL'S TAU BASED ON PARTIALLY ORDERED SYSTEMS

By SOL HABERMAN

University of Minnesota

INTRODUCTION

If a pair of objects a, b are connected by a relation R which is non-reflexive, anti-symmetric and transitive, we say, if aRb, that a *precedes* b. Suppose we have a set of n objects in which R holds between some, not necessarily all, objects. The situation may be represented diagrammatically by a set of points, one to each object, joined by lines between each pair of objects which are connected by R. The relation aRb between any pair exists, if any, only if there is a path along the network joining the corresponding points. We may represent the direction of the relationship by an arrow on the corresponding line, and the relation aRb requires that the arrows along one path, at least between a and b, all go in the same direction. For convenience, in this paper, we shall regard all our directions as going downwards (from the head of the page towards the foot), and it will not be necessary actually to draw the arrows.

On this understanding, any downward path in the diagram corresponds to an ordering of the objects through which it passes. Now suppose that each of the n objects exhibits one of the ranks of each of a set of ranked variables. For example, if there are two variables, A, a dichotomy into male and female (ranked 1, 2), and B, a fourfold division into social class, working, lower-middle, upper-middle and upper (ranked 1, 2, 3, 4), any individual under consideration will bear one of the ranks for each of A and B: the ranking (23), for instance, denoting a female in the upper middle class. In this particular case there are $2 \times 4 = 8$ types of individual and the number n of objects is accordingly 8.

If comparisons are made between neighbouring members differing only in one variable, beginning with (1, 1) and ending with (2, 4) we have what is called a partial ordering. Thus (11), (12), (13), (23), (24) is such an ordering, and so is (11), (21), (22), (23), (24). The situation is illustrated in Fig. 1, both partial orderings corresponding to downward paths on the diagram. Generally, if the objects correspond to rankings of $p_1, p_2, ..., p_r$ a partial ordering will be a set of $p_1 + p_2 + ... + p_r - r + 1$ objects such that any consecutive pair is of the type $(a, b, ..., j, ..., r)$, $(a, b, ..., j+1, ..., r)$. The number n is equal to $p_1 p_2 \cdots p_r$.

There will be $n!$ rankings of the n objects and $\dfrac{(\Sigma p_j - r)!}{(p_1 - 1)!\,(p_2 - 1)! \cdots (p_r - 1)!}$, partial orderings beginning with $(1\,1\,1\,...\,1)$ and ending with $(p_1 p_2 p_3 \cdots p_r)$. For in the process of following through any partial ordering any rank, say the jth, has to make $p_j - 1$ unit moves and these can occur at any point in the $\Sigma(p_j - 1)$ moves. Thus with $p_1 = 2$, $p_2 = 4$ there are four partial orderings:

$$
\left.
\begin{array}{lllll}
(11) & (21) & (22) & (23) & (24) \\
(11) & (12) & (22) & (23) & (24) \\
(11) & (12) & (13) & (23) & (24) \\
(11) & (12) & (13) & (14) & (24)
\end{array}
\right\}
\tag{1}
$$

With $p_1 = 2$, $p_2 = 2$, $p_3 = 2$ there are six partial orderings each comprising four objects, and so on. The totality of partial orderings may be called the set of such orderings.

Consider a ranking of the n objects, e.g. for 2×4 objects,

$$(11) \quad (12) \quad (13) \quad (14) \quad (21) \quad (22) \quad (23) \quad (24). \tag{2}$$

We see that every pair of objects in the partial orderings (1) are in the same order as in the ranking (1). Such a ranking is said to be *consistent* with the set of partial orderings. It is not unique, e.g.

$$(11) \quad (12) \quad (13) \quad (21) \quad (14) \quad (22) \quad (23) \quad (24) \tag{3}$$

has the same property. Any other ranking may be compared with one of the consistent rankings by considering the minimum number of interchanges of adjacent pairs necessary to transform one to the other. More than one consistent ranking may be reached by the same minimum number of interchanges.
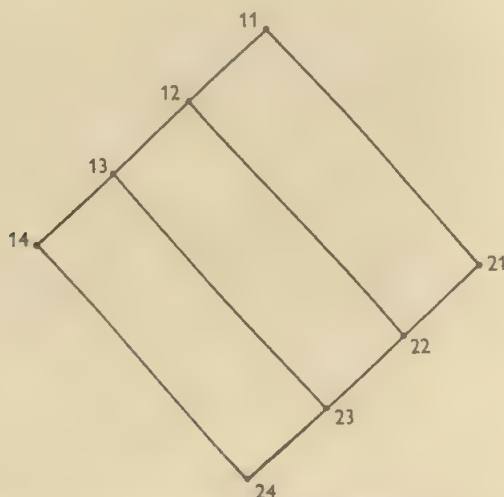
Fig. 1. Partially ordered system for two variables, one dichotomous and the other fourfold.

The problem discussed in this paper may be formulated as follows: A set of partial orderings is given, and a ranking of the objects is observed. Can this be regarded as consistent with the partial orderings? If the observed ranking is not subject to error, of course, this question is simply decided by examining whether it is one of the consistent rankings. We may also, however, allow for error in the observed ranking and regard it as consistent in a probabilistic sense if it is close to a consistent ranking, 'closeness' in this sense being measured by the minimum number of interchanges of adjacent pairs necessary to transform the observed to a consistent ranking.

This number of interchanges will be denoted by s, and our object is to find the frequency distribution of s in the population of $n!$ possible rankings of the objects. Thus, the hypothesis that an observed ranking is consistent with a set of partial orderings will be accepted if s is small; or, equivalently, if the probability of the observed s or a lower value is below some assigned significance level. The procedure is similar to Kendall's use of s to measure the agreement between two rankings (cf. Kendall, 1955). The difference in our case is that

we require the agreement between an observed ranking and one of a number of possible consistent rankings, not a unique one. Kendall's τ is defined in terms of s by

$$\tau = 1 - \frac{4s}{n(n-1)}, \tag{4}$$

but for our purposes it will be sufficient, and more convenient, to work with s itself.

Distribution of s for three dichotomous variables

I consider in the first place three dichotomous variables ($p_1 = p_2 = p_3 = 2$). There are, as noted above, six partial orderings of four. The corresponding diagram is given in Fig. 2. The 8 objects give rise to $8! = 40,320$ rankings. Of these 48 are consistent.
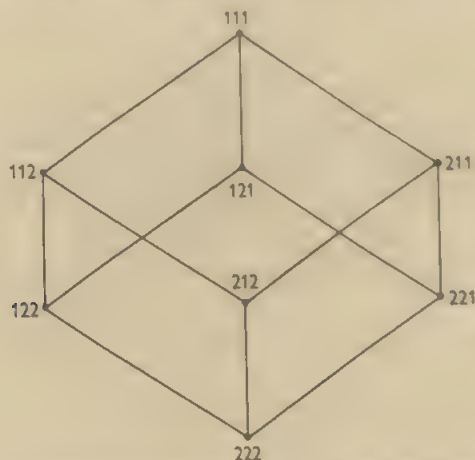
Fig. 2. Partially ordered system for three dichotomous variables.

This last result may be seen as follows: Since (111) comes first and (222) last we need only consider the other six objects. The six partial orderings are (112) (122); (112, 212); (121, 122); (121, 221); (211, 221); (211, 212). These do not impose any conditions on the order of the three (112), (121), (211) or of (122, 212, 221). But any member of the first precedes two members of the second. The first three must then occupy three of the first four places. It will be seen that if there is no overlap between members of the two sets the number of arrangements is $3! \times 3! = 36$. The only possible overlap occurs when one member of the first triad follows a member of the second in the fifth place. This gives rise to 12 further possibilities making 48 in all.

Let (111) be represented by A; each of (211), (121), (112) by B; each of (221), (212), (122) by C; and (222) by D.

In a ranking of eight A and D can occur in 56 different ways, which we can classify into 14 categories as follows:

$$
\begin{array}{ccc}
A & \begin{array}{c} - \\ A \end{array} \quad \begin{array}{c} A \\ - \end{array} & \begin{array}{c} - \\ - \\ A \end{array} \quad \begin{array}{c} - \\ A \\ - \end{array} \quad \begin{array}{c} A \\ - \\ - \end{array} \\
\end{array}
$$

and so on up to

$$
\begin{array}{ccccccc}
- & - & - & - & - & - & A \\
- & - & - & - & - & A & D \\
- & - & - & - & A & D & - \\
- & - & - & A & D & - & - \\
- & - & A & D & - & - & - \\
- & A & D & - & - & - & - \\
A & D & - & - & - & - & - \\
D & - & - & - & - & - & - \\
\end{array} \qquad (7)
$$

The seven categories which follow are a horizontal mirror image of the first seven.

When any given ranking of the three B's and C's is placed in the blank positions it is found that s increases by 1 for each successive category. We can then build up the distribution of s in the $56 \times 720 = 40{,}320$ rankings of 8 when we have the distribution for one case, e.g. the array with A first and D last. It is not easy to ascertain this distribution. There are 720 members and each has to be compared with one of 48 rankings. By writing down the 720 permutations and counting I find the distribution given in Table 1. The distribution for the 40,320 rankings is then obtained as in Table 2.

Table 1. *Distribution of s for the 6 B's and C's*

s	Frequency of s
0	48
1	48
2	84
3	108
4	108
5	108
6	96
7	60
8	36
9	24
Total	720

Some further distributions for the case of two variates are given in Table 3 for the 2×2, 2×3, 2×4 and 3×3 cases.

With four dichotomous variables there are 16! rankings to be considered. (1111) and (2222) occur in 240 ways in a ranking of 16 and, as before, may be treated independently of the remaining 14 combinations in any enumeration for the distribution of s. Since direct enumeration of the distribution of s for the remaining 14 combinations would have been too tedious, an estimate of the form of the distribution was obtained by sampling.

The 14 combinations were written on slips which were placed in a bowl from which drawings were made to obtain a sample of some 2500 rankings. This number was held to be substantial in view of the 240 rankings of 16 associated with each of the 2500. The 240×2500

values of s were weighted according to the theoretical frequencies of certain sub-distributions to which they belonged. When graphed as a frequency polygon, the final estimate of the form of the distribution appeared almost identical with a normal curve. I am unable at this stage to give a more comprehensive account of the distributions of s and partially ordered systems.

Table 2. *Computations used to obtain the distribution of s for the rankings of the eight combinations of three dichotomous variables*

Category	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
1		—	—	—	—								—	24	36	60	96	108	108	108	84	48	48
2			—	—	—							72	108	180	288	324	324	324	252	144	144		
3				—	—					96	144	240	384	432	432	432	336	192	192				
4			—	—	—				120	180	300	480	540	540	540	420	240	240					
5	—	—	—	—	—			144	216	360	576	648	648	648	504	288	288						
6		—	—	—	—		168	252	420	672	756	756	756	588	336	336							
7		—	—	—	168	252	420	672	756	756	756	588	336	336									
8		—	—	—	168	252	420	672	756	756	756	588	336	336									
9		—	—	144	216	360	576	648	648	648	504	288	288										
10	—	—	—	120	180	300	480	540	540	540	420	240	240										
11	—	—	96	144	240	384	432	432	432	336	192	192	—										
12	—	72	108	180	288	324	324	324	252	144	144	—											
13	48	72	120	192	216	216	216	168	96	96	—												
14	24	36	60	96	108	108	108	84	48	48	—												
Totals	24	84	204	420	744	1176	1716	2316	2904	3444	3828	4008	3972	3720	3276	2724	2112	1512	1008	612	324	144	48

Grand total, 40,320

MISSING COMBINATIONS

In the preceding, distributions of s are given for the situation where all of the possible combinations of the ranks of a set of variables are considered. It happens in practice, however, that some combinations do not appear at all or else appear with negligible frequencies. In such cases those combinations cannot be ranked with respect to the others. Since the problem discussed in this paper is to test if an observed ranking of combinations is consistent with a partially ordered system of the same combinations, we have to derive distributions based on partially ordered systems in which those combinations do not appear. This was done in connexion with the derivation of the distribution of s for the case of three dichotomous variables. Table 1 gives the distribution which is based on the partially ordered system in which both (111) and (222) are missing.

When combinations are considered a few at a time, the distributions so formed often fall into classes where within each class all are identical. For instance, when, in the case of three dichotomous variables, the combinations are considered 7 at a time we get either of two forms of distribution of s:

(a) Form F, if one of the B's or C's is missing; and

(b) Form G, if either A or D is missing.

The two forms of distribution are given in Table 4.

Table 3. *Two variable cases*

A, both dichotomous.   B, one dichotomous and the other having three ranks.
C, one dichotomous and the other having four ranks.   D, both having three ranks.

s	Frequencies			
	A	B	C	D
0	2	5	14	42
1	4	15	56	170
2	6	32	150	464
3	6	53	318	1,016
4	4	77	587	1,946
5	2	96	963	3,362
6	—	105	1,448	5,350
7	—	102	2,009	7,942
8	—	88	2,599	11,096
9	—	67	3,151	14,664
10	—	43	3,601	18,442
11	—	24	3,883	22,138
12	—	10	3,958	25,434
13	—	3	3,817	28,010
14	—	—	3,477	29,604
15	—	—	2,990	30,054
16	—	—	2,418	29,306
17	—	—	1,832	27,454
18	—	—	1,289	24,684
19	—	—	839	21,276
20	—	—	495	17,548
21	—	—	260	13,818
22	—	—	115	10,354
23	—	—	41	7,354
24	—	—	9	4,916
25	—	—	1	3,072
26	—	—	—	1,770
27	—	—	—	930
28	—	—	—	432
29	—	—	—	170
30	—	—	—	52
31	—	—	—	10
Totals	24	720	40,320	362,880

Table 4. *Distribution of s for three dichotomous variables with one combination missing*

	Frequencies	
s	Form F (6 distributions)	Form G (2 distributions)
0	16	48
1	48	96
2	104	180
3	184	288
4	284	396
5	396	504
6	500	600
7	572	612
8	604	600
9	588	540
10	528	432
11	436	324
12	324	216
13	220	120
14	132	60
15	68	24
16	28	—
17	8	—
Totals	5040	5040

Table 5. *Distribution of s for the case of two variables, one dichotomous and one fourfold, with one combination missing*

s	Frequencies (two distributions in each form)			
	Form P	Form Q	Form R	Form S
0	14	5	9	7
1	42	20	32	26
2	94	52	75	64
3	168	105	142	125
4	269	182	232	210
5	376	278	338	312
6	485	383	445	419
7	561	480	533	512
8	604	553	587	575
9	594	588	597	596
10	544	578	560	570
11	450	525	484	503
12	344	439	382	407
13	235	337	275	301
14	145	235	178	201
15	74	147	101	119
16	32	80	48	60
17	8	37	18	25
18	1	13	4	7
19	—	3	—	1
Totals	5040	5040	5040	5040

Again, when in the case of two variables where one is dichotomous and the other has four ranked classifications, the combinations are considered 7 at a time and we get four forms of distribution of s:

(a) Form P, if either (11) or (24) is missing;
(b) Form Q, if either (21) or (14) is missing;
(c) Form R, if either (12) or (23) is missing; and
(d) Form S, if either (13) or (22) is missing.

The four forms of distribution are given in Table 5.

The writer is indebted to Prof. F. Stuart Chapin for encouragement and to Prof. Palmer O. Johnson for guidance.

REFERENCE

KENDALL, M. G. (1955). *Rank Correlation Methods*, 2nd ed. London. Chas. Griffin and Co.

ON A CLASS OF SKEW DISTRIBUTION FUNCTIONS

By HERBERT A. SIMON†

Carnegie Institute of Technology

I. INTRODUCTION

It is the purpose of this paper to analyse a class of distribution functions that appears in a wide range of empirical data—particularly data describing sociological, biological and economic phenomena. Its appearance is so frequent, and the phenomena in which it appears so diverse, that one is led to the conjecture that if these phenomena have any property in common it can only be a similarity in the structure of the underlying probability mechanisms. The empirical distributions to which we shall refer specifically are: (A) distributions of words in prose samples by their frequency of occurrence, (B) distributions of scientists by number of papers published, (C) distributions of cities by population, (D) distributions of incomes by size, and (E) distributions of biological genera by number of species.

No one supposes that there is any connexion between horse-kicks suffered by soldiers in the German army and blood cells on a microscope slide other than that the same urn scheme provides a satisfactory abstract model of both phenomena. It is in the same direction that we shall look for an explanation of the observed close similarities among the five classes of distributions listed above.

The observed distributions have the following characteristics in common:

(a) They are J-shaped, or at least highly skewed, with very long upper tails. The tails can generally be approximated closely by a function of the form

$$f(i) = (a/i^k)b^i, \tag{1.1}$$

where a, b, and k are constants; and where b is so close to unity that in first approximation the final factor has a significant effect on $f(i)$ only for very large values of i. Thus, for example, the number of words that occur exactly i times in James Joyce's *Ulysses* is about $a \, i^k$; the number of authors who published exactly i papers in *Econometrica* over a twenty-year period is approximately a/i^k; and so on.

(b) The exponent, k, is greater than 1, and in the cases of word frequencies, publication, and urban populations is very close to 2.‡

(c) In the cases of word frequencies, publications and biological genera, the function (1.1) describes the distribution not merely in the tail but also for small values of i. In these cases the ratio $f(2)/f(1)$ is generally in the neighbourhood of one-third, and almost never reaches one-half; while $f(1)/n$, where $n = \sum_1^\infty f(i)$, is generally in the neighbourhood of one-half.

Property (a) is characteristic of the 'contagious' distributions—for example, the negative binomial as it approaches its limiting form, Fisher's logarithmic series distribution. However, in the case of the negative binomial, k cannot exceed unity (and equals unity only in

‡ See Zipf (1949) for numerous examples of distributions with this property.

the limiting case of the log series); and if the distribution has a long tail, so that the convergence factor, b, is close to unity. $f(2)/f(1)$ cannot be less than one-half. Hence the negative binomial and Fisher's log series distributions do not provide a satisfactory fit for data possessing property (a) together with either (b) or (c).†

It is well known that the negative binomial and the log series distributions can be obtained as the stationary solutions of certain stochastic processes. For example, J. H. Darwin (1953) derives these from birth and death processes, with appropriate assumptions as to the birth- and death-rates and the initial conditions. In this paper we shall show that stochastic processes closely similar to those yielding the negative binomial or log series distributions lead to a class of functions having the three properties enumerated above. This class of functions is given by

$$f(i) = AB(i, \rho+1), \tag{1.2}$$

where A and ρ are constants, and $B(i, \rho+1)$ is the Beta function of $i, \rho+1$:

$$B(i, \rho+1) = \int_0^1 \lambda^{i-1}(1-\lambda)^\rho \, d\lambda = \frac{\Gamma(i)\,\Gamma(\rho+1)}{\Gamma(i+\rho+1)} \quad (0<i; \; 0<\rho<\infty). \tag{1.3}$$

Now it is a well-known property of the Gamma function (Titchmarsh, 1939, p. 58) that as $i \to \infty$, and for any constant, k,

$$\frac{\Gamma(i)}{\Gamma(i+k)} \sim i^{-k}. \tag{1.4}$$

Hence, from (1.3), we have, as $i \to \infty$:

$$f(i) \sim \Gamma(\rho+1)\, i^{-(\rho+1)}. \tag{1.5}$$

Therefore, the distribution (1.2) approximates the distribution (1.1) in the tail (more precisely, through the range in which the convergence factor of the latter is close to one). Further, if ρ is positive, k will be greater than 1, as required by (b); and if ρ is equal to 1, k will be equal to 2. It is easy to see that in the latter case we will have

$$f(i) = \frac{1}{i(i+1)}, \quad \sum_{i=1}^\infty f(i) = 1, \tag{1.6}$$

so that $f(2)/f(1) = \frac{1}{3}$; and $f(1)/n = \frac{1}{2}$, as required by (c).

In the remainder of this paper I propose: (a) to describe a stochastic process that leads to the stationary distribution (1.2); (b) to discuss some generalizations of this process; and (c) to construct hypotheses as to why the empirical phenomena mentioned above can be represented, approximately, by processes of this general kind. Before proceeding, I should like to mention two earlier derivations, one of (1.2), the other of (1.1), that I have been able to discover in the literature.

Some thirty years ago, G. Udny Yule (1924) constructed a probability model, with (1.2) as its limiting distribution, to explain the distribution of biological genera by numbers of species. He also derived a modified form of (1.2), replacing the complete Beta-function of (1.3) by the incomplete Beta-function with upper limit of integration $\alpha < 1$. (This modification has the same effect as the introduction of the convergence factor, b^i, in (1.1)—it causes a more rapid decrease in $f(i)$ for very large values of i; cf. also Darwin (1953, p. 378).) It seems highly appropriate to call the distribution (1.2) the Yule distribution.

† The contrasting characteristics of distributions for which the log series provides a satisfactory fit and those, under consideration here, for which it does not are illustrated by examples (i) and (ii), respectively, in Good (1953).

Because Yule's paper predated the modern theory of stochastic processes his derivation was necessarily more involved than the one we shall employ here. Moreover, while the assumptions he required are plausible for the particular biological problem he treated, the corresponding assumptions applied to the four other phenomena we have mentioned appear much less plausible. Our derivation requires substantially weaker assumptions than Yule's about the underlying probability mechanism.

More recently D. G. Champernowne (1953) has contructed a stochastic model of income distribution that leads to (1·1) and to generalizations of that function. Since the points of similarity between his model and the one under discussion here are not entirely obvious at a first examination, I shall consider their relation in a later section of this paper.

II. The Stochastic Model

For ease of exposition, the model will be described in terms of word frequencies. In a later section, alternative interpretations will be provided. Our present interest is in the kind of stochastic process that would lead to (1·2).

Consider a book that is being written, and that has reached a length of k words. We designate by $f(i, k)$ the number of *different* words that have occurred exactly i times in the first k words. That is, if there are 407 different words that have occurred exactly once each, then $f(1, k) = 407$.

Assumption I. The probability that the $(k+1)$-st word is a word that has already appeared exactly i times is proportional to $if(i, k)$—that is, to the total number of occurrences of all the words that have appeared exactly i times.

Note that this assumption is much weaker than the assumption (I'): that the probability a *particular* word occur next be proportional to the number of its previous occurrences. Assumption (I') implies (I), but the converse is not true. Hence we leave open the possibility that, among all words that have appeared i times the probability of recurrence of some may be much higher than of others.

Assumption II. There is a constant probability, α, that the $(k+1)$-st word be a new word— a word that has not occurred in the first k words.

Assumptions (I) and (II) describe a stochastic process, in which the probability that a particular word will be the next one written depends on what words have been written previously. If this process correctly describes the selection of words, then the words in a book cannot be regarded as a random sample drawn from a population with a prior distribution. The reasonableness of the former, as compared with the latter type of explanation of the observed distributions, will be discussed in §IV.

From (I), it follows that

$$\mathscr{E}\{f(i, k+1)\} - f(i, k) = K(k)\{(i-1)f(i-1, k) - if(i, k)\} \quad (i = 2, \dots, k+1), \qquad (2\cdot1)$$

for if the $(k+1)$st word is one that has previously occurred $(i-1)$ times, $f(i, k+1)$ will be increased over $f(i, k)$, and the probability of this, by assumption (I), is proportional to $(i-1)f(i-1, k)$; if the $(k+1)$st word is one that previously occurred i times, $f(i, k+1)$ will be decreased, and the probability of this, by assumption (I), is proportional to $if(i, k)$; while in all other cases, $f(i, k+1) = f(i, k)$.

From (I) and (II) we obtain similarly

$$\mathscr{E}\{f(1, k+1)\} - f(1, k) = \alpha - K(k)f(1, k) \quad (0 < \alpha < 1). \qquad (2\cdot2)$$

Since we will be concerned throughout with 'steady-state' distributions (as defined by equation (2·8) below), we replace the expected values in (2·1) and (2·2) by the actual frequencies. (Alternatively, we might replace frequencies on the right-hand side of the equation by probabilities.) That is, we write, instead of (2·1) and (2·2),

$$f(i, k+1) - f(i, k) = K(k)\{(i-1)f(i-1, k) - if(i, k)\} \quad (i = 2, \ldots, k+1), \tag{2·3}$$

$$f(1, k+1) - f(1, k) = \alpha - K(k)f(1, k), \tag{2·4}$$

where the f's now represent expected values.

Now, we wish to evaluate the factor of proportionality $K(k)$. Since $K(k)if(i, k)$ is the probability that the $(k+1)$st word is one that previously occurred i times, we must have

$$\sum_{i=1}^{k} K(k)if(i, k) = K(k) \sum_{i=1}^{k} if(i, k) = 1 - \alpha. \tag{2·5}$$

But $\sum_{i=1}^{k} if(i, k)$ is the total number of words up to the kth, hence

$$\sum_{i=1}^{k} if(i, k) = k, \tag{2·6}$$

and

$$K(k) = \frac{1-\alpha}{k}. \tag{2·7}$$

Substituting (2·7) in (2·3) and (2·4), we could solve these differential equations explicitly. We can avail ourselves, however, of a simpler—though non-rigorous—method for discovering the solutions, and can then test their correctness by substitution in the original equations. Consider the 'steady-state' distribution in the following sense. We assume

$$\frac{f(i, k+1)}{f(i, k)} = \frac{k+1}{k} \quad \text{for all } i \text{ and } k; \tag{2·8}$$

so that all the frequencies grow proportionately with k, and hence maintain the same relative size. (Since we must have $f(i, k) = 0$ for $i > k$, equation (2·8) cannot hold exactly for all i and k. But as explained above, we are concerned at the moment with heuristic rather than proof.)

From (2·8) it follows that

$$\frac{f(i, k)}{f(i-1, k)} = \frac{f(i, k+1)}{f(i-1, k+1)} = \beta(i), \tag{2·9}$$

where $\beta(i)$ does not involve k. Hence, the *relative* frequencies, which we will designate by $f^*(i)$, are independent of k. Substituting (2·7), (2·8) and (2·9) in (2·3), we get

$$\left(\frac{k+1}{k} - 1\right)f(i, k) = \frac{(1-\alpha)}{k}\left\{\frac{(i-1)}{\beta(i)} - i\right\}f(i, k). \tag{2·10}$$

Cancelling the common factor, and solving β for (i), we obtain

$$\beta(i) = \frac{(1-\alpha)(i-1)}{1+(1-\alpha)i} = \frac{f^*(i)}{f^*(i-1)} \quad (i = 2, \ldots, k). \tag{2·11}$$

For convenience, we introduce

$$\rho = \frac{1}{1-\alpha} \quad (1 < \rho < \infty). \tag{2·12}$$

Since $f^*(i) = \beta(i)f^*(i-1) = \beta(i).\beta(i-1)\ldots\beta(2)f^*(1)$, we obtain from (2.11) and (2.12)

$$f^*(i) = \frac{(i-1)(i-2)\ldots 2.1}{(i+\rho)(i+\rho-1)\ldots(1+\rho)}f^*(1) = \frac{\Gamma(i)\Gamma(\rho+1)}{\Gamma(i+\rho+1)}f^*(1) = B(i,\rho+1)f^*(1) \quad (i=2,\ldots,k) \quad (2.13)$$

The second relation follows from the fact that

$$\Gamma(i+\rho+1) = (i+\rho)\Gamma(i+\rho) = (i+\rho)(i+\rho-1)\ldots(1+\rho)\Gamma(\rho+1) \quad (2.14)$$

But (2.13) is identical with (1.2) if we take $A = f^*(1)$.

That (2.13) is in fact a solution of (2.3) can be verified by direct substitution. Moreover, it is in the following sense a stable solution. Suppose that (2.11) is *not* satisfied. Whatever be the values of the $f(i,k)$ for a given k, we may write without loss of generality

$$\frac{f(i,k)}{f(i-1,k)} = \frac{(1-\alpha)(i-1)}{(1-\alpha)i+1+\epsilon(i,k)}, \quad (2.15)$$

where $\epsilon(i,k)$ is some function of i and k. If we now divide both sides of (2.3) by $f(i,k)$ and substitute (2.15) in the right-hand side of the resulting equation, we obtain after simplification

$$\frac{f(i,k+1)}{f(i,k)} = \frac{k+1+\epsilon(i,k)}{k}. \quad (2.16)$$

Hence the ratio of $f(i,k+1)$ to $f(i,k)$ will be greater than $(k+1)/k$ if $\epsilon(i,k)$ is positive, and less than $(k+1)/k$ if $\epsilon(i,k)$ is negative. Since new words are introduced at a constant rate, $\sum_1^k f(i,k)$ must be proportional to k; therefore, by (2.16), we will have

$$\sum_{i=1}^{k+1} f(i,k+1) - \frac{k+1}{k}\sum_{i=1}^{k+1} f(i,k) = \frac{1}{k}\sum_{i=1}^{k}\epsilon(i,k)f(i,k) = 0. \quad (2.17)$$

We may interpret the three equations, (2.15) (2.17), as follows. In an average sense, the frequencies will grow proportionately with k. If a particular frequency is 'too large' compared with the next lower frequency ($\epsilon(i,k)$ negative in (2.15)), it will grow at a rate slower than the average; if it is 'too small' ($\epsilon(i,k)$ positive), it will grow more rapidly than the average.

It remains to be shown that $f^*(i) = B(i,\rho+1)f^*(1)$ is a proper distribution function. In particular, we require that $\sum_{i=1}^{k} iB(i,\rho+1)$ converge as $k\to\infty$. Now, it is well known that $\sum_{i=1}^{\infty} i^{-a}$ converges for every $a > 1$. But by (1.4),

$$\sum_{i=1}^{\infty} iB(i,\rho+1) \sim i.i^{-(\rho+1)} = i^{-\rho}. \quad (2.18)$$

Hence, by the usual ratio comparison test, $\sum_{i=1}^{\infty} iB(i,\rho+1)$ converges for $\rho > 1$, as required.

From the definition of α the total number, n_k, of *different* words will be αk; while the total number of word occurrences is k. That is

$$n_k = \sum_{i=1}^{k} f(i,k) = \alpha k = \alpha\sum_{i=1}^{k} if(i,k). \quad (2.19)$$

Returning to (2·4), and using (2·8), we get

$$\left(\frac{k+1}{k}-1\right)f^*(1) = \alpha - \frac{1-\alpha}{k}f^*(1),$$ (2·20)

whence

$$f^*(1) = \frac{k\alpha}{2-\alpha} = \frac{n_k}{2-\alpha}.$$ (2·21)

From (2·12) and (2·21), and by successive application of (2·11), we can compute the values of $\rho, f^*(1)/n_k, f^*(2)/n_k, f^*(3)/n_k$, etc., for given values of α (Table 1).

Table 1

α	ρ	$f^*(1)/n_k$	$f^*(2)/n_k$	$f^*(3)/n_k$
0·0	1	0·500	0·167	0·083
0·1	1·11	0·527	0·169	0·082
0·2	1·25	0·556	0·171	0·080
0·3	1·43	0·588	0·171	0·077
0·5	2·00	0·667	0·167	0·067
0·7	3·33	0·769	0·144	0·046
0·9	10·00	0·909	0·076	0·012

Thus far we have considered the case where α, the rate at which new words are introduced, is independent of k. We can easily generalize to the case where α is a function of k by making the appropriate substitution in (2·4). The equations can then be solved directly, but the method employed to obtain a 'steady-state' distribution is not applicable, since it is not easy to define what is meant by the steady state in this more general case. We will content ourselves with some approximate results for two special cases. These special cases will give us insight as to how a distribution function may arise which, for small values of i, can be approximated by (1·2), with $0 < \rho < 1$.

Case I. Suppose the system to be in the steady state described by (2·13) with $k = k_0$, and that the flow of *new* words suddenly ceases, so that $\alpha(k) = 0$ for $k > k_0$. We will now have $K(k) = 1/k$ for $k > k_0$, and (2·4) becomes

$$f(1, k+1) = \left(1-\frac{1}{k}\right)f(1, k) = \frac{k-1}{k}f(1, k).$$ (2·22)

We define

$$\gamma(i) = \frac{f(i, k+1)}{f(i, k)} \quad (i = 2, \dots, k+1).$$ (2·23)

Since no new words are being introduced, we must have

$$n_k = f(1, k) + \sum_{i=2}^{k} f(i, k) = f(1, k+1) + \sum_{i=2}^{k} f(i, k+1)$$

$$= \frac{(k-1)}{k}f(1, k) + \sum_{i=2}^{k} \gamma(i)f(i, k),$$ (2·24)

whence

$$\frac{\sum_{i=2}^{k}[\gamma(i)-1]f(i, k)}{\sum_{i=2}^{k} f(i, k)} = \frac{1}{k}\frac{f(1, k)}{\sum_{i=2}^{k} f(i, k)}.$$ (2·25)

Let us define next
$$\beta(i) = \frac{f(i,k)}{f(i-1,k)} = \frac{(i-1)}{(1+\rho_i)} \tag{2.26}$$

(where we suppose that ρ_i changes only slowly with k). Instead of (2.3), we have

$$f(i,k+1) - f(i,k) = \frac{1}{k}[(i-1)f(i-1,k) - if(i,k)]. \tag{2.27}$$

Substituting (2.23) and (2.26) in this, we get

$$\gamma(i) - 1 = \frac{1}{k}[(i+\rho_i) - i], \tag{2.28}$$

whence
$$\rho_i = k(\gamma(i) - 1), \tag{2.29}$$

and
$$\bar{\rho} = \frac{\sum\limits_{i=2}^{k} k(\gamma(i)-1)f(i,k)}{\sum\limits_{i=2}^{k} f(i,k)} = \frac{f(1,k)}{\sum\limits_{i=2}^{k} f(i,k)} = \frac{f(1,k)}{n_k - f(1,k)}. \tag{2.30}$$

Define
$$p_1 = f(1,k)/n_k. \tag{2.31}$$

Then
$$\bar{\rho} = \frac{p_1}{1-p_1} \quad \text{and} \quad 0 < \bar{\rho} < \infty. \tag{2.32}$$

Proceeding heuristically, we can see that after α becomes zero, $f(1,k)$ will begin to decrease with k, and the value of ρ_i will be larger the larger is i. For small values of i, we will have $\rho(i) < \bar{\rho}$, and for large values, $\rho(i) > \bar{\rho}$. However, the tail of the distribution will be affected only slowly by the change in α. Hence, we may suppose that $\lim\limits_{i \to k_0} \rho(i) = \rho_0$, where ρ_0 is $\rho(k_0)$. On the other hand, since the weighted average in (2.29) is heavily influenced by the large frequencies for small values of i, ρ_i will be only slightly less than $\bar{\rho}$. Hence we may expect the distribution to take the form of a slightly curved line on a double-log scale, with a slope of $-(\bar{\rho}+1)$ at the lower end, and a slope of $-(\rho_0+1)$ at the upper end. If $\rho_0 > 2$, then $\Sigma if(i,k)$ will converge. An example of such a distribution will be given in § IV.

Case II. A second approximate solution can be obtained if we assume that α decreases with k, but very slowly. By definition, we have $\alpha = dn_k / dk = n'$. The condition for a steady state (all frequencies increasing proportionately) is now

$$f(i,k+1) = [1 + (n'/n_k)]f(i,k). \tag{2.33}$$

Substituting as before, (2.7) and (2.33) in (2.3), we again obtain (2.13), where ρ is now given by
$$\rho = \frac{n'k}{n_k} \frac{1}{(1-n')}. \tag{2.34}$$

The slope obtained in the derivation for constant α has now been multiplied by the factor $(n'k)/n_k$, which for monotonically decreasing α is less than one. Hence, the effect of a decrease in the rate of introduction of new words is to lengthen the tail of the distribution, as was also true in case I. If the new value of ρ is less than one, we do not have a proper distribution function (see equation (2.18)), hence the equation can hold only for small and moderate values of i, and there must be a curve (on a logarithmic scale) in the tail of the distribution.

III. An alternative formulation of the process

There are some alternative ways for deriving the relation (2·13). One of these will be useful to us when we come, in the next section, to a more specific discussion of word frequencies and frequencies of publications. Moreover, this derivation avoids the difficulties we have encountered in the definition of 'steady state'.

Equation (2·10) may be written

$$0 = (1-x)[(i-1)f^*(i-1) - if^*(i)] - f^*(i) \quad (i = 2, ..., k), \tag{3·1}$$

where we have again written $f^*(i)$ for $f(i, k)$.

Similarly, from (2·4), we obtain

$$0 = 1 - (1-\alpha)f^*(1) - f^*(1). \tag{3·2}$$

These two equations may be interpreted as follows. We consider a sequence of k words. We add words to the sequence in accordance with assumptions (I) and (II) of § II, *but we drop words from the sequence at the same average rate*, so that the length of the sequence remains k. The method according to which we drop words is the following:

Assumption III. If one representative of a particular word is dropped, then all representatives of that word are dropped, and the probability that the next word dropped be one with exactly i representatives is $f^*(i)$.

This assumption would be approximately satisfied, for example, if the representatives of each word, instead of being distributed randomly through the sequence, were closely 'bunched'. This possibility is consistent with assumption (I).

Equation (3·1), in our new interpretation, may be regarded as the steady-state equilibrium of the stochastic process defined by

$$f(i, m+1) - f(i, m) = (1-x)[(i-1)f(i-1, m) - if(i, m)] - f(i, m), \tag{3·3}$$

where m is now not the total number of words (which remains a constant, k), but the number of additions to (and withdrawals from) an initial arbitrary sequence of k words. Since the k of this process, unlike that of § II, remains constant, the ordinary proofs of the existence of a unique steady-state solution will apply (see Feller, 1950, p. 373), and we avoid the troublesome questions of rigour that confronted us in § II.

The solution of (3·1) and (3·2) is, of course, again given by

$$\frac{f^*(i)}{f^*(i-1)} = \frac{(1-\alpha)(i-1)}{1+(1-\alpha)i}. \tag{2·11}$$

If we were to replace the last terms of (3·1) and of (3·2), respectively, by terms corresponding to the usual form of the death process, we would have (cf. Darwin, 1953, p. 375; and Kendall, 1948)

$$0 = (1-x)[(i-1)f^*(i-1) - if^*(i)] - [if^*(i) - (i+1)f^*(i+1)] \quad (i = 2, ..., k-1). \tag{3·4}$$

$$0 = 1 - (1-\alpha)f^*(1) - [f^*(1) - 2f^*(2)]. \tag{3·5}$$

The solution of this system of equations is

$$\frac{f^*(i)}{f^*(i-1)} = \frac{(1-\alpha)(i-1)}{i}, \tag{3·6}$$

which is Fisher's logarithmic series distribution.

Since the log series distribution is a limiting case of the negative binomial, we may ask whether there is a distribution that stands in the same relation to the latter as (2·11) stands in relation to (3·6). We can obtain such a distribution by a modification of the birth process in (3·1). We assume now that the birth-rate is the sum of two components—one proportional to $if(i)$, the other proportional to $f(i)$. In place of (3·1) we have

$$0 = \frac{(1-\alpha)k}{k+c}[(i-1+c)f^*(i-1)-(i+c)f^*(i)] \quad f^*(i) \quad (c \text{ a constant}). \tag{3·7}$$

the solution of which is

$$\frac{f^*(i)}{f^*(i-1)} = \frac{\lambda(i-1+c)}{\lambda(i+c)+1} = \frac{(i-1+c)}{(i+c+1/\lambda)}, \tag{3·8}$$

where

$$\lambda = k(1-\alpha)/(k+c).$$

A rather remarkable property of (3·8) is that in the tail it still has the limiting form (1·1) with $b = 1$. Hence for α and c small, this generalized Yule distribution will still possess the three properties listed in the introduction. The fact that a reasonably wide range of variation in the assumptions underlying the stochastic model does not alter greatly the form of the distribution adds plausibility to the use of such stochastic processes to explain the observed distributions. Our next task is to consider these explanations in more detail.

IV. THE EMPIRICAL DISTRIBUTIONS

In this section I shall try to give theoretical justifications for the observed fit of the Yule distribution to a number of different sets of empirical data.

A. *Word frequencies*

A substantial number of word counts have been made, in English and in other languages (see Hanley, 1937; Thorndike, 1937; Yule, 1944; Zipf, 1949; and Good, 1953). Equation (1·6) provides a good fit to almost all of them. When the more general function, (1·2), is used, the estimated value of ρ is always close to 1. When a convergence factor, b^i, is introduced to account for the deficiency in frequencies for very large values of i, the estimated value of b is also very close to 1. Good (1953), for instance, applies (1·6) multiplied by a convergence factor to the Eldridge count, and obtains $b = 0.999667$.

These regularities are the more surprising in that the various counts refer to a quite heterogeneous set of objects. In the Yule and Thorndike counts, inflected forms are counted with the root word; in most of the other counts each form is regarded as a distinct word. The Yule counts include only nouns; the others, all parts of speech. The Dewey, Eldridge and Thorndike counts are composite—compiled from a large number of separate writings; most of the others are based on a single piece of continuous prose. I would regard this heterogeneity as further evidence that the explanation is to be sought in a probability mechanism, rather than in more specific properties of language; but at the same time, the heterogeneity complicates the task of specifying the probability mechanism in detail. I shall avoid questions of 'fine structure'—which would require an expertness in linguistics that I do not possess—and confine myself to three broad problems: (1) the distribution of word frequencies in the whole historical sequence of words that constitutes a language; (2) the distribution of word frequencies in a continuous piece of prose; (3) the distribution of word frequencies in a sample of prose assembled from composite sources.

(1) For obvious reasons, we do not have any empirical data on the cumulated word frequencies for a whole language. On *a priori* grounds, it does not appear unreasonable to postulate that these frequencies are determined by a process like that described in §II. The parameter α is then the rate at which neologisms appear in the language as a fraction of all word occurrences —and hence α can be assumed to be very close to zero.

(2) The process of §II might also describe the growth of a continuous piece of prose—for example, Joyce's *Ulysses*. But there are some serious objections to this hypothesis. An author writes not only by processes of *association*—i.e. sampling earlier segments of the word sequence—but also by processes of *imitation*—i.e. sampling segments of word sequences from other works he has written, from works of other authors, and, of course, from sequences he has heard. The model of §II apparently allows only for association, and excludes imitation.

The word frequencies in *Ulysses* provide obvious evidence of the importance of both processes. The fact that the proper noun 'Bloom' occurs 926 times and ranks 30th in frequency must be attributed to association. If Joyce had named his hero 'Smith', that noun, instead of 'Bloom', would have ranked 30th. On the other hand, 'they', which occurs 1010 times in *Ulysses* and ranks 27th, has very nearly the same rank – the 28th— in the Dewey count. In fact, of the 100 most frequent words in *Ulysses*, 78 are among the top 100 in the Dewey count. This similarity in ranking of 'common' words argues for imitation rather than association. Even for the common words, however, the variations in frequency from one count to another are far too great to be explained as fluctuations resulting from random sampling from a common population of words. The imitative process must involve stratified sampling, and imitation must be compounded with association.

It is worth emphasizing again at this point that assumption (I) does not require that the choice of the next word from among those previously written be completely random. Suppose, for example, that a writer were to assign to each page he has already written a number, p_j, $\Sigma p_j = (1-\alpha)$, the size of p_j varying with the 'affinity' of the subject discussed on the jth page to the subject next to be discussed. If his next word were selected by a stratified sampling of the previous pages, with probability p_j for each page, then equation (2·1) would generally be satisfied. For although individual words would be distributed unevenly through the preceding pages, the totality of words having a given frequency, i, in all the previous pages taken together would be distributed almost evenly through these pages. Hence, the various frequency strata would have proportionate probabilities of being sampled, for most choices of the p_j. This is all that is required for equation (2·1). This same comment applies to the assumption we shall subsequently make regarding imitative sampling from other works.

Let us now reconsider the problem of a piece of continuous prose. Since both the processes of association and imitation are involved, the sequence that is counted is to be regarded as a 'slice', of length k, of the entire sequence of words in the language, or of the entire sequence written by the author. Hence the word count is better described by the stochastic process of §III than by the process of §II.

In determining the probability that a word selected in such a sequence be one that has occurred exactly i times, we must consider separately the process of imitation and association. Assume that, on the average, a fraction, β, of the words added is selected by imitation, and the remaining fraction, $(1-\beta)$, by association. Since no new words can be introduced by association, the joint probability that the next word will be selected by association and will be a word that has already occurred i times is $(1-\beta) i f(i, k) / k$.

The words selected by imitation present a more difficult problem, and we shall have to content ourselves with a reasonable assumption that has no rigorous justification. On the average, a word that has occurred i times will have a chance less than ik of being the next one chosen by imitation, because in the sequence that is being sampled there are words that have not yet been chosen at all, and because with progressive change of subject, different strata of the language will be sampled. Since words with large i will generally be 'common' words, fairly uniformly distributed through all strata of the language, the deficiency may be expected to be proportionately greater for small i than for large i. As a rough, but reasonable, approximation let us assume that: the joint probability that the next word will be selected by imitation and will be a word that has already occurred i times is $\beta(i-c)f(i,k)/k$, where $0 < c < 1$. (Our result would not be essentially altered if we wrote $c(i)$ instead of c, provided only that $c(i)$ does not vary a great deal.)

Adding the two joint probabilities—for association and imitation, respectively—we find that the total probability that the next word be one that has occurred i times is $(i-\beta c)f(i,k)/k$. By summing this probability over i and subtracting from 1, we find that the probability that the next word be a new word is $\beta c(n_k, k)$.

If the method of dropping words from the sequence satisfies assumption (III) of §III, we set the difference between the birth-rate and the death-rate equal to zero, and obtain the steady-state equation

$$0 = (i - c\beta - 1)f^*(i-1) - (i - c\beta)f^*(i) - f^*(i), \qquad (4\cdot1)$$

which has as its solution
$$\frac{f^*(i)}{f^*(i-1)} = \frac{(i - c\beta - 1)}{(i - c\beta + 1)}. \qquad (4\cdot2)$$

Again, we obtain a distribution with the required properties.

(3) The distribution of word frequencies in a sample of prose assembled from composite sources can be explained along the same general lines. Again, we may regard the sample as a 'slice' from a longer sequence, but we might expect the parameters c and β to be somewhat larger than in a comparable piece of continuous prose. The qualification 'comparable' is important, for c may be expected to be smaller for homogeneous prose using a limited vocabulary of common words than for prose with a large vocabulary and treating of a variety of subjects. Hence c might well be larger for the continuous *Ulysses* count than for the Eldridge count, which is drawn from newspaper sources. Indeed, the empirical evidence suggests that this is the case.

There is no point in elaborating the explanation further. What has been shown is that the observed frequencies can be fitted by distributions derived from probability assumptions that are not without plausibility.

A very different and very ingenious explanation of the observed word-frequency data has been advanced recently by Dr Benoit Mandelbrot (1953). His derivation rests on the assumption that the frequencies are determined so as to maximize the number of bits of information, in the sense of Shannon, transmitted per symbol. There are several reasons why I prefer an explanation that employs averaging rather than maximizing assumptions. First, an assumption that word usage satisfies some criterion of efficiency appears to be much stronger than the probability assumptions required here. Secondly, numerous doubts, which I share, have been expressed as to the relevance of Shannon's information measure for the measurement of semantic information.

Before leaving the subject of word frequencies, it may be of interest to look, at some of the empirical data. Good (1953, pp. 257-60), has obtained good fits to the Eldridge count, and to one of Yule's counts by the use of equation (1·6). Table 2, summarizes a few of the data on two word counts, and compares the actual frequencies, $f(1)$, $f(2)$, and $f(3)$ with the frequencies estimated from equation (1·3). The actual values of k and n_k are used to estimate $\alpha = n_k/k$, and (2·11) and (2·21) to obtain the expected frequencies. In both cases the observed value of n_k/k leads to an estimate of ρ in the neighbourhood of 1·1 to 1·2. An empirical fit to the whole distribution of a function of the form $f(i) = Ka^{i}i^{-\rho-1}$ gives an estimated value of ρ, in both cases, of about one in reasonable agreement with the first estimate. A good fit to both the *Ulysses* and the Eldridge counts can also be obtained from (4·2), with c equal to about 0·2 in the former case, and close to zero in the latter.

In the case of Thorndike's count of $4\frac{1}{2}$ million words in children's books (Thorndike, 1937) we may assume that the supply of new words was virtually exhausted before the end of the count. In his count $f(1)$ is substantially below $0·5n_k$ (about $0·34n_k$), as we would expect under these circumstances (see case I of § II). Thorndike estimated the empirical value of our $\bar{\rho}$ at 0·45, which is entirely consistent with the observed value of $0·34n_k$ for $f(1)$. For, by (2·32), $p_1 = \bar{\rho}/(\bar{\rho}+1) = 0·31$.

<div align="center">Table 2</div>

Word count	$\alpha = \dfrac{n_k}{k}$	$f(1)$		$f(2)$		$f(3)$	
		Actual	Estimate	Actual	Estimate	Actual	Estimate
Ulysses (Hanley, 1937)	0·115	16,432	15,850	4,776	4,870	2,194	2,220
Eldridge (Good, 1953)	0·136	2,976	3,220	1,079	977	516	400

B. *Scientific publications*

At least four sets of data are available on the number, $f(i)$, of authors contributing a given number, i, of papers each to a journal or journals (Davis, 1941; Leavens, 1953). These are counts of (a) papers written by members of the Chicago Section of the American Mathematical Society over a 25-year period; (b) papers listed in *Chemical Abstracts* (under A and B) over 10 years; (c) papers referred to in a history of physics; and (d) papers and abstracts in *Econometrica* over a 20-year period.

We may postulate a mechanism like that of § III, equation (3·1). The authorship of the next paper to appear is 'selected' by stratified sampling from the strata of authors who have previously published 1, 2, ..., papers, the probability for each stratum being proportional to $if(i)$. Again, the probabilities for individual authors need not be proportional to i, but only the probabilities for the aggregates of authors with the same i. For example (as in the case of words), the probability for a particular author may be higher if he has published recently than if he has not. The gradual retirement of authors corresponds to assumption (III).

A comparison of the actual frequencies, for i from 1 to 10, with the estimated frequencies, derived from (2·11) and (2·21), is shown in Table 3. The fit is reasonably good, when it is remembered that only one parameter is available for adjustment. However, it should be

...ground that the estimated frequencies tend to be too high for $i = 1$, 2 and too low for $i = 3$, ... In three of the four cases, they are again too high for the two tails of the distributions. A further refinement of the model is apparently needed to remove these discrepancies.

Table 3. *Number of persons contributing*

No. of contributions	Chaung-Math. Sci.		Chem. Physics		Econometrica		Psychometrika	
	Actual[†]	Estimate	Actual[‡]	Estimate	Actual[†]	Estimate	Actual[‡]	Estimate
1	133	—	3,991	4,000	784	824	436	453
2	43	46	1,059	1,160	204	217	107	119
3	24	23	493	522	127	94	61	94
4	12	14	287	288	50	50	40	27
5	11	10	184	179	33	30	14	16
6	14	7	131	120	28	20	23	11
7	5	5	113	86	19	14	6	7
8	3	4	85	64	19	10	11	5
9	9 (13)	3 (10)	64	49	6 (32)	8 (24)	1 (12)	4 (12)
10	1	3	66	38	7	6	0	3
11 or more	23	30	419	335	48	52	22	26
Estimated α	0§		0·30		0·39		0·41	
Estimated ρ	0·916§		1·43		1·64		1·69	
k	1,124		22,939		3,496		1,759	
n_k	278		6,891		1,325		721	
ρ'	1·07							

† Davis (1941).
‡ Leavens (1953).
§ $\rho = \bar{\rho}$ estimated in this case from (2·31) to (2·32).

C. *City sizes*

It has been observed, for every U.S. Census since the early nineteenth century, and for most other Western countries as well, that if $F(i)$ is the number of cities of population greater than i, then

$$F(i) \sim Ai^{-\rho}, \tag{4.3}$$

where ρ is close to 1 (see Zipf, 1949, chs. 9, 10).

Again, we would expect such a distribution if the underlying mechanism were one describable by equations like (2·3) and (2·4). Such a mechanism is not hard to conceive. First, equation (2·3) would hold if the growth of population were due solely to the net excess of births over deaths, and if this net growth were proportional to present population. This assumption is certainly satisfied at least roughly. Moreover, it need not hold for each city, but only for the aggregate of cities in each population band. Finally, the equation would still be satisfied if there were net migration to or from cities of particular regions, provided the net addition or loss of population of individual cities *within any region* was proportional to city size. That is, even if all California cities were growing, and all New England cities declining, the equation would hold provided the percentage growth or decline in each area were uncorrelated with city size.

In the case of cities, equation (4·3) could only be expected to hold down to some minimum city size—say, 5000 or 10,000. The constant α would then be interpreted as the fraction of the total population growth in cities above the minimum size that is accounted for by the new cities that reach that size.

D. *Income distribution*

Vilfredo Pareto is generally credited with the discovery that if personal incomes are ranked by size, the number of persons, $F(i)$, whose incomes exceed i can be approximated closely, for the upper ranges of income, by equation (4·3) with ρ usually in the neighbourhood of 1·5 (Davis, 1941; Champernowne, 1953). Hence, the income distributions bear a family resemblance in their upper ranges to those we have already considered, although the parameter, ρ, is substantially larger than 1—its characteristic value in the case of word frequencies and city size distributions.

A stochastic mechanism similar to those described in §III would again produce steady-state distributions closely resembling the observed ones. We picture the stream of income as a sequence of dollars allocated probabilistically to the recipients. If the total annual income of all persons above some specified minimum income is k dollars, the segment of this sequence running from the mth to the $(m+k)$th dollar is the income for the year beginning at time m. We assume that the probability that the next dollar will be allotted to some person with an annual income of i dollars is proportional to $(i+c)f(i)$, with c positive but small. This represents a modification of assumption (I) that decreases the proportion of the total stream going to persons of high income relative to the proportion going to persons with incomes close to the minimum. We assume that a fraction of the dollars is assigned to new persons—i.e. persons reaching the minimum income to which the assumptions apply (assumption (II)). We assume that there is considerable variance among persons within each income class in the probability of receiving additional income, so that the rate at which dollars are dropped from any income class as m increases satisfies assumption (III). Then we obtain again equation (3·8), which now holds for i greater than the minimum income. For large i, this distribution has the required properties with $1/\lambda = \rho$.

The same result has been reached by D. G. Champernowne (1953), following a somewhat different route. He divides income recipients at time t_1 into classes of equal proportionate width. That is, if i_m is the minimum income considered, then the first class contains persons with incomes between i_m and ri_m, the second class, persons with incomes between ri_m and $r^2 i_m$, and so on. Next he introduces transition probabilities p_{gh}, that a person who is in class g at time t_1 will be in class h at time t_2. He assumes that p_{gh} is a function only of $(g-h)$. Now, by his definition of the income classes, the average income of persons in class g will be about $r^{(g-h)}$ times the average income of persons in class h. Hence, the expected income at t_2 of a person who was in class g at t_1 will be

$$\sum_h p_{gh} i_h = \sum_h p_{(g-h)} r^{(g-h)} i_g = \alpha i_g \quad (\alpha \text{ a constant}), \tag{4·4}$$

where i_g is the average income in class g. Prof. Champernowne assumes explicitly that $\alpha < 1$. From this it is clear that his model satisfies our assumptions (I) (in its original form) and (II). Further, since he assumes a substantial variance in income expectations among persons in a given class, our assumption (III) is also approximately satisfied. Hence, in spite of the surface differences between his model and those developed here, the underlying structure is the same.

E. *Biological species*

We conclude this very incomplete list of phenomena exhibiting the Yule distribution by mentioning the example originally analysed by Yule himself (1924). It was discovered by Willis that the number, $f(i)$, of genera of plants having i species each was distributed approximately according to (4·3), with $\rho < 1$. Yule explained these data by a probability model in which the probability, s, of a specific mutation occurring in a particular genus during a short time interval was proportional to the number of species in the genus; while the probability, r, of a generic mutation during the same interval was proportional to the number of genera. Starting at t_0 with a single genus of one species, he computed the distribution $f(i, t)$ for $t_1, t_2, \ldots$, and found the limit as $t \to \infty$. This limiting distribution corresponds to (2·13) with $\rho = r/s$. Yule observed that for $r < s$ (as required to fit the empirical data), this was not a proper distribution function, and obtained the approximate distribution for $t = T$. His procedure was equivalent to replacing the complete Beta function in (2·13) by the incomplete Beta function, taking as the upper limit of integration an appropriate function of T.

If, in the process of §II, we define k as the total number of different species and $f(i, k)$ as the number of genera with exactly i species, we see that our k is a monotonic increasing function of Yule's t (specifically, $k = e^{st}$). Making the appropriate transformation of variables, we find that Yule's assumption with respect to the rate of specific mutation corresponds to our assumption (I′) (and hence is considerably stronger than the assumption we employed in §II). Making the same transformation of variables with respect to his assumption of a constant rate of generic mutation, we find that $n_k = e^{rt}$. We can then compute $\alpha(k)$ (which will now vary with k) by taking the derivative of n_k with respect to k. We obtain

$$\alpha(k) = r\, e^{(r-s)t}/s. \tag{4·5}$$

If we substitute these values in equation (2·34) of case II, where we assumed slowly changing α, we find in the limit, as $t \to \infty$, $\rho = r/s$, as required. Hence, we see that the process of §II is essentially the same as the one treated by Yule.

It is interesting and a little surprising that when Yule, some twenty years after this discovery, examined the statistics of vocabulary, he did not employ this model to account for the observed distributions of word frequencies. Indeed, in his fascinating book on *The Statistical Study of Literary Vocabulary* (1944) he nowhere refers to his earlier paper on biological distributions.

V. CONCLUSION

This paper discusses a number of related stochastic processes that lead to a class of highly skewed distributions (the Yule distribution) possessing characteristic properties that distinguish them from such well-known functions as the negative binomial and Fisher's logarithmic series. In §I, the distinctive properties of the Yule distribution were described. In §§II and III several stochastic processes were examined from which this distribution can be derived. In §IV, a number of empirical distributions that can be approximated closely by the Yule distribution were discussed, and mechanisms postulated to explain why they are determined by this particular kind of stochastic process. In the same section, the derivations of §§II and III were compared with models previously proposed by Yule (1924) and Champernowne (1953) to account for the data on biological species and on incomes, respectively.

The probability assumptions we need for the derivations are relatively weak, and of the same order of generality as those commonly employed in deriving other distribution functions—the normal, Poisson, geometric and negative binomial. Hence, the frequency with which the Yule distribution occurs in nature—particularly in social phenomena should occasion no great surprise. This does not imply that all occurrences of this empirical distribution are to be explained by the process discussed here. To the extent that other mechanisms can be shown also to lead to the same distribution, its common occurrence is the less surprising. Conversely, the mere fact that particular data conform to the Yule distribution and can be given a plausible interpretation in terms of the stochastic model proposed here tells little about the underlying phenomena beyond what is contained in assumptions (I) through (III).

REFERENCES

CHAMPERNOWNE, D. G. (1953). A model of income distribution. *Econ. J.* **63**, 318.

DARWIN, J. H. (1953). Population differences between species growing according to simple birth and death processes. *Biometrika*, **40**, 370.

DAVIS, HAROLD T. (1941). *The Analysis of Economic Time Series.* Principia Press.

FELLER, WILLIAM (1950). *An Introduction to Probability Theory and Its Applications*, vol. **1**. Wiley.

GOOD, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, **40**, 237.

HANLEY, MILES L. (1937). *Word Index to James Joyce's Ulysses.* University of Wisconsin Press.

KENDALL, DAVID G. (1948). On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. *Biometrika*, **35**, 6.

LEAVENS, DICKSON H. (1953). *Econometrica*, **21**, 630.

MANDELBROT, BENOIT (1953). An informational theory of the statistical structure of language. In *Communication Theory* (ed. by Willis Jackson). Butterworths.

THORNDIKE, EDWARD L. (1937). On the number of words of any given frequency of use. *Psychol. Rec.* **1**, 399.

TITCHMARSH, E. C. (1939). *The Theory of Functions*, 2nd ed. Oxford University Press.

YULE, G. UDNY (1924). A mathematical theory of evolution, based on the conclusions of Dr J. C. Willis, F.R.S. *Phil. Trans.* B, **213**, 21.

YULE, G. UDNY (1944). *The Statistical Study of Literary Vocabulary.* Cambridge University Press.

ZIPF, GEORGE KINGSLEY (1949). *Human Behavior and the Principle of Least Effort.* Addison-Wesley Press.

SIMULTANEOUS TESTS OF LINEAR HYPOTHESES

By M. N. GHOSH

University of North Carolina and University of Calcutta

1. INTRODUCTION

A very common situation in the analysis of variance of survey data, where the investigator is not able to put his experimental data in the framework of a design, planned in advance, is that the estimates of the parameters are correlated in various ways and the analysis of variance becomes cumbersome. Even for the analysis of variance of a two-way classification with unequal class frequencies one meets with this difficulty. We shall be concerned here with the analysis of such data and the test of significance for groups of parameters representing different aspects of the problem. We consider below a concrete example, in which standing height, sexual maturity characters and certain blood chemicals, like haemoglobin, ascorbic acid, carotene and alkaline phosphates were measured for a number of girls between the ages 9 and 14 years.[*] For the purpose of analysis the girls were divided into twenty-five classes according to sexual maturity ratings from indices of breast (B_i) and pubic hair development (Ph_j), each of these being a 5-point rating.

The following model is assumed:

$$y_{k(i,j)} = \mu_{ij} + \beta_1 x_{1,k} + \beta_2 x_{2,k} + \beta_3 x_{3,k} + \gamma z_k + \epsilon_k, \tag{1.1}$$

where $(y_k, x_{1,k}, x_{2,k}, x_{3,k}, z_k)$ $(k = 1, 2, ..., N)$ are sample individuals. $y_{k(i,j)}$ represents the height increment of the kth girl in the maturity class $(B_i Ph_j)$. $x_{1,k}, x_{2,k}, x_{3,k}$ are nutritional variables haemoglobin, ascorbic acid and carotene respectively, and z_k is alkaline phosphatase. We thus consider three groups of parameters $\{\mu_{ij}\}$ $(i, j = 1, ..., 5)$. $\{\beta_1, \beta_2, \beta_3\}$ and $\{\gamma\}$. The hypotheses corresponding to these three groups of parameters are

$$\left.\begin{array}{ll} H_1: & \mu_{ij} = \mu \quad (i, j = 1, 2, ..., 5), \\ H_2: & \beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \\ H_3: & \gamma = 0. \end{array}\right\} \tag{1.2}$$

Rejection of the hypothesis H_1 would imply that the height increment depends upon sexual maturity, the rejection of H_2 would imply that the blood chemicals, which depend upon nutritional status, affect height increment, and the rejection of H_3 would imply that alkaline phosphatase affects height increment.

The hypotheses H_1, H_2, H_3 may be tested by the analysis of variance method, but these tests are not independent, because of the non-orthogonality of the groups of estimates as well as for the fact that we have to use the same estimate of error variance. We shall introduce the notion of quasi-independent tests of multiple hypotheses, which is appropriate in such situations.

---

[*] The data were collected by Dr Hughes Bryan, Professor of Nutrition, School of Public Health, University of North Carolina.

2. Quasi-independent tests of multiple hypotheses

For any test of significance we consider the first and second kinds of error. Let the hypotheses tested be

$$H_1: \theta_1 = 0, \quad H_2: \theta_2 = 0, \quad H_3: \theta_3 = 0. \tag{2.1}$$

The tests T_1, T_2, T_3 of H_1, H_2, H_3 will be called quasi-independent when

$$\left. \begin{aligned} \text{Prob}_{T_1} \{\text{acceptance } H_1 \,|\, \theta_1 \neq 0, \theta_2, \theta_3\} &= \text{Prob}_{T_1} \{\text{acceptance } H_1 \,|\, \theta_1 \neq 0, \theta_2 = 0, \theta_3 = 0\} \\ \text{Prob}_{T_1} \{\text{rejection } \quad H_1 \,|\, \theta_1 = 0, \theta_2, \theta_3\} &= \text{Prob}_{T_1} \{\text{rejection } \quad H_1 \,|\, \theta_1 = 0, \theta_2 = 0, \theta_3 = 0\} \end{aligned} \right\} \tag{2.2}$$

hold for all values of θ_2 and θ_3, i.e. whether H_2 and H_3 are true or not; and similar equations hold for T_2 and T_3, where $\text{Prob}_T \{ \ \}$ denotes the probability of the statement in parentheses for the test procedure T_1, etc. Thus for quasi-independent tests of hypotheses H_1, H_2 and H_3 the first and second kinds of error for each hypothesis do not depend on the parameters of the other hypotheses.

Two tests of hypotheses H_1 and H_2 with the critical regions C_1 and C_2 are independent if the following holds for all θ_1 and θ_2:

$$\text{Prob}\{X \subset C_1, X \subset C_2 \,|\, \theta_1, \theta_2\} = \text{Prob}\{X \subset C_1 \,|\, \theta_1\}, \text{Prob}\{X \subset C_2 \,|\, \theta_2\},$$

X being the observed sample.

It will be seen later that the usual analysis of variance tests for multiple hypotheses are quasi-independent, and we most often do not need independent tests.

3. Control of errors in the simultaneous tests of hypotheses

There may be different points of view for assigning significance levels, in the case of simultaneous tests of hypotheses. In certain situations, where the decisions regarding the hypotheses $H_1, H_2, \ldots$, etc., are unrelated, it is proper to consider the significance level of each hypothesis individually at 5 or 1 % level (say). But when the decisions have a joint import, it is proper to consider the first kind of error of the simultaneous test of H_1 and H_2, as the rejection of at least one of the hypotheses, when all of them are in fact true.

The significance level of the simultaneous test is defined as the probability of rejecting at least one of the hypotheses, when all of them are true.

Suppose in the test of a new variety of crop against a common variety used as a control we are interested in two different characters and the new variety may be considered desirable when it is superior in either of these characters. Since one would be interested to know which particular character in the new variety is superior to the control, a simultaneous test should be done. In this case, the significance level, as some sort of measure of the amount of caution implied in the test, should naturally be that of the simultaneous test, since one would like to insure, at a certain level, against declaring the new variety superior, when it is actually no better than the control.

A similar situation exists in quality control for acceptance of material, when a number of characters are examined and the material is rejected when it is not up to the mark with respect to any of these characters. It may be useful, in this case, to determine the particular deficiency in any of these characters, and so we must use a simultaneous test, and the significance level of the simultaneous test should be used to insure against too frequent rejections. Of course, we pay for this by widening confidence intervals for the parameters measuring these characters.

The need for a safety device like the simultaneous significance level would be even more apparent for the agricultural or quality control example stated above when the number of characters is large, in which case the chance of declaring a new variety different from the common variety may be much larger than 5%. However, the control of the first kind of error in a simultaneous test is achieved only at the expense of increasing the second kind of error for individual characters. In any particular problem, whether the point of view of individual or simultaneous significance level should be adopted depends, roughly, on how much *a priori* weight one attaches to the alternative hypotheses, either from theoretical expectations or from considerations of cost in replacing the old variety by the new variety. If this cost is small, e.g. if the varieties are grown only on an experimental scale, one would be relatively free to decide on either of the varieties as superior and the individual significance level would be the proper one. However, with varieties established in agricultural practice, any statistical decision in favour of a new variety would involve large expenditures and the statistician should take an attitude of caution. The simultaneous significance level takes account of this attitude and is thus appropriate in such cases.

In the analysis of variance problem considered before, the conventional statistical procedure of (a) testing for the overall SS for fitting constants, (b) testing separately for the SS of H_1, adjusting for H_2 at 5% level, etc., would give quasi-independent tests at individual levels of significance 5%. But the problem of obtaining the simultaneous significance level in this case is mathematically a very intractable one, and we shall find an upper bound for the significance level. This would give a control of the first kind of error. The operating characteristic would, of course, be of the same nature as for the usual analysis of variance of multiple hypotheses, since the test procedure is essentially the same, except that the exact significance level is not known but only its upper bound.

The notion of simultaneous level of significance has already been considered in various ways and languages by Scheffé (1953), Tukey (1952) and Nandi (1951), and the practical implications have been thoroughly discussed by Tukey in a mimeographed report.

4. Simultaneous tests and significance levels

Consider the linear model

$$E(y_i) = a_{i1}p_1 + \ldots + a_{im}p_m \quad (i = 1, 2, \ldots, N), \tag{4.1}$$

y_i being independent normal variables with unknown variance σ^2, and $p_1, \ldots, p_m$ are unknown parameters. Let rank $(a_{ij}) = N_0$ and let $\pi_1 \ldots \pi_R$ be estimable linear functions of the parameters $p_1, \ldots, p_m$

$$\pi_\kappa = l_{\kappa 1}p_1 + l_{\kappa 2}p_2 + \ldots + l_{\kappa m}p_m \quad (\kappa = 1, 2, \ldots, R), \tag{4.2}$$

such that the coefficient vectors $(l_{\kappa 1}, \ldots, l_{\kappa m})$ form a vector space of rank $R \leqslant N_0$. We consider the following linear hypotheses,

$$\left. \begin{array}{l} H_1: \quad \pi_1 = 0, \ \ldots, \ \pi_{\kappa_1} = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ H_s: \quad \pi_{\kappa_1 + \ldots + \kappa_{s-1} + 1} = 0, \ \ldots, \ \pi_{\kappa_1 + \ldots + \kappa_s} = 0. \end{array} \right\} \tag{4.3}$$

Let $Y_1, \ldots, Y_{\kappa_1}; Y_{\kappa_1+1}, \ldots, Y_{\kappa_1+\kappa_2}, \ldots, Y_{\kappa_1+\ldots+\kappa_s}$ be the best linear estimates of these parameters, obtained by the method of least squares. We shall sometimes denote the coefficient vectors

of these linear functions by the same symbol, so that we have the alternative notation for the linear function $Y_i = (Y_i, y)$, where (Y_i, y) is the scalar product of the vector Y_i and the vector y. From Markoff's theorem we have an independent estimate of error variance, S_e^2, with n_e d.f., which is independent of the parameters $p_1, \ldots, p_m$. Quasi-independent tests of the hypotheses H_n can be made by considering the linear functions

$$(Y_{b_{n-1}+1}, y) \ldots (Y_{b_n}, y) \quad (b_n = \kappa_1 + \ldots + \kappa_n),$$

whose expectations are $\pi_{b_{n-1}+1}, \ldots, \pi_{b_n}$. Let $U_{b_{n-1}+1}, \ldots, U_{b_n}$ be orthonormal vectors forming a basis of the vector space formed by $Y_{b_{n-1}+1}, \ldots, Y_{b_n}$. Then (U_i, y) is a linear form in (Y_j, y) $(i, j = b_{n-1}+1, \ldots, b_n)$ and

$$E\{(U_i, y)\} = \sum_{j=b_{n-1}+1}^{b_n} \alpha_j E(Y_j, y) = \sum_{j=b_{n-1}+1}^{b_n} \alpha_j \pi_j \quad \text{from} \quad E(Y_j, y) = \pi_j. \tag{4.4}$$

On the hypothesis H_n, $\chi_n^2/\sigma^2 = \sum_{j=b_{n-1}+1}^{b_n} [(U_j, y) - E(U_j, y)]^2$ has a χ^2-distribution with κ_n d.f. and $\chi_n^2/(\kappa_n S_e^2)$ has an F-distribution with d.f. (r, n_e). The second kind of error for the test of H_n depends only upon the parameters $\pi_{b_{n-1}+1}, \ldots, \pi_{b_n}$.

We now consider sets $M_1, M_2, \ldots$ of vectors $Y_1, \ldots, Y_{\kappa_1}; Y_{\kappa_1+1}, \ldots, Y_{\kappa_1+\kappa_2}; \ldots$ belonging to hypotheses $H_1, \ldots, H_s$, so that all vectors belonging to the same hypothesis H_i belong to the same set M_l. Vectors belonging to different sets are orthogonal, and if the vectors belonging to two hypotheses H_i and H_j are orthogonal they belong to different sets. These sets we shall call orthogonal sets and hypotheses belonging to two different sets, orthogonal hypotheses. We shall consider different cases according as an orthogonal set consists of a single hypothesis H_i or more.

Case 1. Let all hypotheses $H_1, \ldots, H_s$ be orthogonal so that each orthogonal set M_l consists of a single hypothesis. Let $U_1, \ldots, U_{\kappa_1}; U_{\kappa_1+1}, \ldots, U_{\kappa_1+\kappa_2}; \ldots, U_{\kappa_1+\ldots+\kappa_s}$ be orthonormal systems of vectors in the spaces determined by $Y_1, \ldots, Y_{\kappa_1}; Y_{\kappa_1+1}, \ldots, Y_{\kappa_1+\kappa_2}; \ldots, Y_{\kappa_1+\ldots+\kappa_s}$ respectively. Let

$$E\{(U_i, y)\} = \Phi_i.$$

We shall consider for simplicity $s = 3$. The joint distribution of

$$\chi_1^2 = \sum_{i=1}^{\kappa_1} \{(U_i, y) - \Phi_i\}^2, \quad \chi_2^2 = \sum_{i=\kappa_1+1}^{\kappa_1+\kappa_2} \{(U_i, y) - \Phi_i\}^2, \quad \chi_3^2 = \sum_{i=\kappa_1+\kappa_2+1}^{\kappa_1+\kappa_2+\kappa_3} \{(U_i, y) - \Phi_i\}^2. \tag{4.5}$$

and S_e^2 is given by

$$\text{const. exp}\left\{-\frac{1}{2\sigma^2}[\chi_1^2 + \chi_2^2 + \chi_3^2 + n_e S_e^2]\right\} (\chi_1^2)^{\frac{1}{2}(\kappa_1-2)} (\chi_2^2)^{\frac{1}{2}(\kappa_2-2)} (\chi_3^2)^{\frac{1}{2}(\kappa_3-2)} (S_e^2)^{\frac{1}{2}(n_e-2)} d\chi_1^2 d\chi_2^2 d\chi_3^2 dS_e^2.$$

Put

$$\frac{\chi_1^2}{n_e S_e^2} = G_1, \quad \frac{\chi_2^2}{n_e S_e^2} = G_2, \quad \frac{\chi_3^2}{n_e S_e^2} = G_3.$$

Making the transformation, integration for S_e^2 gives the distribution of G_1, G_2, G_3 as

$$C(\kappa_1, \kappa_2, \kappa_3; n_e) \frac{G_1^{\frac{1}{2}(\kappa_1-2)} G_2^{\frac{1}{2}(\kappa_2-2)} G_3^{\frac{1}{2}(\kappa_3-2)}}{(1+G_1+G_2+G_3)^{\frac{1}{2}(\kappa_1+\kappa_2+\kappa_3+n_e)}} dG_1 dG_2 dG_3. \tag{4.6}$$

We now consider the region defined by $G_1 < \lambda_1$, $G_2 < \lambda_2$, $G_3 < \lambda_3$ which has the probability $P(\lambda_1, \lambda_2, \lambda_3)$, say. If $P(\lambda_1, \lambda_2, \lambda_3) = 1 - \alpha$, we have the system of simultaneous confidence regions $C_1, C_2, \ldots$ for the sets of parameters with the confidence coefficient $1 - \alpha$, as

$$
\begin{aligned}
C_1: & \quad \pi_1, \ldots, \pi_{\kappa_1}: \quad \sum_{i=1}^{\kappa_1} \{(U_i, y) - \Phi_i\}^2 < \lambda_1 n_e S_e^2, \\
C_2': & \quad \pi_{\kappa_1+1}, \ldots, \pi_{\kappa_1+\kappa_2}: \quad \sum_{i=\kappa_1+1}^{\kappa_1+\kappa_2} \{(U_i, y) - \Phi_i\}^2 < \lambda_2 n_e S_e^2, \\
C_3': & \quad \pi_{\kappa_1+\kappa_2+1}, \ldots, \pi_{\kappa_1+\kappa_2+\kappa_3}: \quad \sum_{i=\kappa_1+\kappa_2+1}^{\kappa_1+\kappa_2+\kappa_3} \{(U_i, y) - \Phi_i\}^2 < \lambda_3 n_e S_e^2.
\end{aligned} \tag{4.7}
$$

The best choice of $\lambda_1, \lambda_2, \lambda_3$, i.e. to give the smallest confidence regions, is not known and needs further investigation. From considerations of degrees of freedom we may consider $\lambda_1/\kappa_1 = \lambda_2/\kappa_2 = \lambda_3/\kappa_3 = \lambda$.

5. EVALUATION OF SIGNIFICANCE LEVEL

To determine λ from (4.6) we have

$$
C(\kappa_1, \kappa_2, \kappa_3; n_e) \int_0^{\kappa_1\lambda} dG_1 \int_0^{\kappa_2\lambda} dG_2 \int_0^{\kappa_3\lambda} dG_3 \frac{G_1^{\frac{1}{2}(\kappa_1-2)} G_2^{\frac{1}{2}(\kappa_2-2)} G_3^{\frac{1}{2}(\kappa_3-2)}}{(1 + G_1 + G_2 + G_3)^{\frac{1}{2}(\kappa_1+\kappa_2+\kappa_3+n_e)}} = 1 - \alpha, \tag{5.1}
$$

where α is the significance level. Put $t_1 = G_1$, $t_2 = G_2(1 + G_1)$, $t_3 = G_3(1 + G_1 + G_2)$, then the above becomes

$$
C(\kappa_1, \kappa_2, \kappa_3; n_e) \int_0^{\kappa_1\lambda} \frac{t_1^{\frac{1}{2}(\kappa_1-2)} dt_1}{(1+t_1)^{\frac{1}{2}(\kappa_1+n_e)}} \left[\int_0^{\kappa_2\lambda/(1+t_1)} \frac{t_2^{\frac{1}{2}(\kappa_2-2)} dt_2}{(1+t_2)^{\frac{1}{2}(\kappa_1+\kappa_2+n_e)}} \int_0^{\kappa_3\lambda.[(1+t_1)(1+t_2)]} \frac{t_3^{\frac{1}{2}(\kappa_3-2)} dt_3}{(1+t_3)^{\frac{1}{2}(\kappa_1+\kappa_2+\kappa_3+n_e)}} dt_3 \right]. \tag{5.2}
$$

which can be evaluated by successive numerical integration and using Pearson's Tables of Incomplete B-functions.

For any confidence coefficient $1 - \alpha$, the calculation of λ depends upon calculations of iterated integrals, and tables have to be prepared for these. One difficulty of preparing such tables is that the integrals depend upon the parameters κ_1, κ_2, κ_3; n_e, etc., and thus unless the number of parameters can be reduced, construction of tables would be difficult. We may, however, get an upper bound for the significance level α, from an inequality of Kimball (1951) given below:

$$
\Pr\{|G_1| \leqslant \lambda_1, |G_2| \leqslant \lambda_2, |G_3| \leqslant \lambda_3\} \geqslant \Pr\{|G_1| \leqslant \lambda_1\} \Pr\{|G_2| \leqslant \lambda_2\} \Pr\{|G_3| \leqslant \lambda_3\}. \tag{5.3}
$$

Here we choose $\lambda_1, \lambda_2, \lambda_3$ so that

$$
\Pr\{|G_1| \leqslant \lambda_1\} = \Pr\{|G_2| \leqslant \lambda_2\} = \Pr\{|G_3| \leqslant \lambda_3\}. \tag{5.4}
$$

Thus if $\Pr\{|G_1| \leqslant \lambda_1, |G_2| \leqslant \lambda_2, |G_3| \leqslant \lambda_3\} = 0.95$ we have to make $\Pr\{|G_1| \leqslant \lambda_1\} = 0.983$. Now $n_e G_1$ has an F-distribution with κ_1 and n_e d.f. so that $\lambda_1 = \frac{\kappa_1}{n_e} F_\alpha(\kappa_1, n_e)$, where $F_\alpha(\kappa, n_e)$ is the $\alpha\%$ point of the F-distribution with d.f. (κ_1, n_e). We thus have

$$
\Pr\left\{|G_1| \leqslant \frac{\kappa_1}{n_e} F_{\alpha'}(\kappa_1, n_e), |G_2| \leqslant \frac{\kappa_2}{n_e} F_{\alpha'}(\kappa_2, n_e), |G_3| \leqslant \frac{\kappa_3}{n_e} F_{\alpha'}(\kappa_3, n_e)\right\} \geqslant (1-\alpha')^3 = 1 - \alpha \quad \text{say.}
$$
$$\tag{5.5}$$

Thus

$$
\alpha' = 1 - (1-\alpha)^{\frac{1}{3}}.
$$

Case II. We now consider a set M_j. which consists of more than one group of linear functions. say it contains linear functions corresponding to the hypotheses H_1 and H_2. We shall call this a compound set. In this case the linear functions belonging to H_1 and H_2 are non-orthogonal. We shall show that there is no non-singular linear transformation by which these linear functions can be transformed into mutually orthogonal sets corresponding to hypotheses H_1 and H_2 respectively. Thus the basic inequality (5·3) is not directly applicable.

Let $U_1, ..., U_{\kappa_1}; U_{\kappa_1+1}, ..., U_{\kappa_1+\kappa_2}$ be an orthonormal basis of the vector space formed by $Y_1, ..., Y_{\kappa_1}; Y_{\kappa_1+1}, ..., Y_{\kappa_1+\kappa_2}$, so that $U_1, ..., U_{\kappa_1}$ is a basis of the vector space formed by $Y_1, ..., Y_{\kappa_1}$. As before, let $E\{(U_i, y)\} = \Phi_i$, then Φ_i are linear functions of $\pi_1, ..., \pi_{\kappa_1}; \pi_{\kappa_1+1}, ..., \pi_{\kappa_1+\kappa_2}$. Let $\tilde{Y}_i$ be the normalized vector corresponding to Y_i, i.e. $\tilde{Y}_i = \dfrac{Y_i}{|Y_i|}$, where $|Y_i|$ is the norm of the vector Y_i and let $\tilde{\pi}_i = \dfrac{\pi_i}{|Y_i|}$. The relation between U-vectors and Y-vectors is expressed by

$$\begin{pmatrix} \tilde{Y}_I \\ \tilde{Y}_{II} \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ \beta & \gamma \end{pmatrix} \begin{pmatrix} U_I \\ U_{II} \end{pmatrix}, \tag{5·6}$$

where $\alpha = (\alpha_{i,j})$ is a $(\kappa_1 \times \kappa_1)$ matrix. $\gamma = (\gamma_{ij})$ is a $(\kappa_2 \times \kappa_2)$ matrix, $\beta = (\beta_{ij})$ is a $(\kappa_2 \times \kappa_1)$ matrix, $\tilde{Y}_I$ is a $(1 \times \kappa_1)$ matrix with components $\tilde{Y}_1, ..., \tilde{Y}_{\kappa_1}$, $\tilde{Y}_{II}$ is a $(1 \times \kappa_2)$ matrix with components $\tilde{Y}_{\kappa_1+1}, ..., \tilde{Y}_{\kappa_1+\kappa_2}$ and similarly U_I is a $(1 \times \kappa_1)$ matrix with components $U_1, ..., U_{\kappa_1}$. etc. From the choice of the basis $U_1, ..., U_{\kappa_1}; U_{\kappa_1+1}, ..., U_{\kappa_1+\kappa_2}$, it is clear that the matrices α and γ are non-singular and $\beta \neq 0$. since the vectors $Y_1, ..., Y_{\kappa_1}$ and $Y_{\kappa_1+1}, ..., Y_{\kappa_1+\kappa_2}$ are non-orthogonal. Inverting the equation (5·6) we get

$$\begin{pmatrix} U_I \\ U_{II} \end{pmatrix} = \begin{pmatrix} \alpha^{-1}, & 0 \\ \gamma^{-1}\beta\alpha^{-1}, & \gamma^{-1} \end{pmatrix} \begin{pmatrix} \tilde{Y}_I \\ \tilde{Y}_{II} \end{pmatrix}. \tag{5·7}$$

Replacing the vectors $\tilde{Y}_i$ by $\tilde{\pi}_i$ and U_i by Φ_i, the corresponding relation must hold between the Φ's and the $\tilde{\pi}$'s. Hence $\Phi_1, ..., \Phi_{\kappa_1}$ depend upon $\tilde{\pi}_1, ..., \tilde{\pi}_{\kappa_1}$ alone but $\Phi_{\kappa_1+1}, ..., \Phi_{\kappa_1+\kappa_2}$ cannot be expressed solely in terms of $\tilde{\pi}_{\kappa_1+1}, ..., \tilde{\pi}_{\kappa_1+\kappa_2}$, since $(\gamma^{-1}\beta\alpha^{-1}) \neq 0$. Thus the sets of linear functions $U_1, ..., U_{\kappa_1}$ and $U_{\kappa_1+1}, ..., U_{\kappa_1+\kappa_2}$ do not provide quasi-independent tests of H_1 and H_2. We shall see later in II (c) that this can be done when the transformation is singular, i.e. when we chose a suitable subset of $\kappa_2 - \kappa_1$ vectors from the vectors $U_{\kappa_1+1}, ..., U_{\kappa_1+\kappa_2}$.

We shall also consider two other methods of setting inequalities to reduce the problem to orthogonal sets. The essence of these methods is to cover a complicated region C by a sphere S and using the probability of the sphere for the region C as an upper bound.

Method (a). *Method of simultaneous confidence intervals for all contrasts*

This method is due to Scheffé (1953). Tukey (1952) and Roy & Bose (1953). It can be used for the simultaneous test but it also gives the confidence intervals for all linear functions of the parameters $\pi_1, ..., \pi_{\kappa_1}; \pi_{\kappa_1+1}, ..., \pi_{\kappa_1+\kappa_2}$ simultaneously with a given confidence coefficient. In equation (5·6) since both U's and $\tilde{Y}$'s are normalized vectors

$$\begin{aligned} \sum_{i=1}^{\kappa_1} \alpha_{ni}^2 = 1 \quad (n = 1, ..., \kappa_1), \\ \sum_{i=1}^{\kappa_1} \beta_{ni}^2 + \sum_{i=1}^{\kappa_2} \gamma_{ni}^2 = 1 \quad (n = 1, ..., \kappa_2). \end{aligned} \tag{5·8}$$

Also

$$\begin{aligned} (\tilde{Y}_n, y) = \sum_{i=1}^{\kappa_1} \alpha_{ni}(U_i, y) \quad (n = 1, 2, ..., \kappa_1), \\ (\tilde{Y}_{\kappa_1+n}, y) = \sum_{i=1}^{\kappa_1} \beta_{ni}(U_i, y) + \sum_{i=1}^{\kappa_2} \gamma_{ni}(U_{\kappa_1+n}, y) \quad (n = 1, 2, ..., \kappa_2). \end{aligned} \tag{5·9}$$

and similar equations hold between π's and Φ's. From Schwartz's inequality

$$[(\tilde{Y}_n, y) - \tilde{\pi}_n]^2 \leqslant \sum_{j=1}^{\kappa_1 + \kappa_2} [(U_j, y) - \Phi_j]^2 \quad (n = 1, 2, \ldots, \kappa_1 + \kappa_2).$$

Hence $\quad \Pr\{|(\tilde{Y}_n; y) - \tilde{\pi}_n| \leqslant \delta \ (n = 1, \ldots, \kappa_1 + \kappa_2)\} \geqslant \Pr\left\{\sum_{j=1}^{\kappa_1 + \kappa_2} (U_j, y) - \Phi_j]^2 \leqslant \delta^2\right\}.$ (5·10)

For the set M_l we now use $\chi_l^2 = \sum_{j=1}^{\kappa_1 + \kappa_2} [(U_j, y) - \Phi_j]^2$ in (4·7). This method has the advantage that if π is a linear function of $\tilde{\pi}_i$'s, then at the same time, with the same confidence coefficient, we have the confidence interval for π,

$$\begin{aligned} \pi &= \alpha_1 \tilde{\pi}_1 + \ldots + \alpha_{\kappa_1 + \kappa_2} \tilde{\pi}_{\kappa_1 + \kappa_2}, \\ &\alpha_1 \tilde{Y}_1 + \ldots + \alpha_{\kappa_1 + \kappa_2} \tilde{Y}_{\kappa_1 + \kappa_2} \pm \delta \sqrt{(\Sigma \alpha_i^2)}. \end{aligned}$$ (5·11)

Method (b)

From the equation (5·6) we have $\quad \tilde{Y}_{II} = \beta U_I - \gamma U_{II}.$ (5·12)

Since the matrix γ is non-singular

$$\gamma^{-1} \tilde{Y}_{II} = \gamma^{-1} \beta U_I + U_{II}.$$ (5·13)

Let $\gamma^{-1} \tilde{Y}_{II} = z_{II}$, which is a column vector with components $z_{\kappa_1 + 1}, \ldots, z_{\kappa_1 + \kappa_2}$, and let $\tilde{z}_{\kappa_1 + 1}, \ldots, \tilde{z}_{\kappa_1 + \kappa_2}$ be the corresponding normalized vectors, then

$$\begin{aligned} (\tilde{z}_{\kappa_1 + 1}, y) &= \beta'_{1,1}(U_1, y) + \ldots + \beta'_{1,\kappa_1}(U_{\kappa_1}, y) + \gamma'_1(U_{\kappa_1 + 1}, y), \\ (\tilde{z}_{\kappa_1 + \kappa_2}, y) &= \beta'_{\kappa_2,1}(U_1, y) + \ldots + \beta'_{\kappa_2,\kappa_1}(U_{\kappa_1}, y) + \gamma'_2(U_{\kappa_1 + \kappa_2}, y). \end{aligned}$$ (5·14)

so that $\sum_{i=1}^{\kappa_1} \beta'^2_{ni} + \gamma'^2_n = 1$ for $n = \kappa_1 + 1, \ldots, \kappa_1 + \kappa_2$, and similar equations hold between the expected values $\psi_j = E(\tilde{z}_j, y)$ and $\Phi_i = E(U_i, y)$. Hence from Schwartz's inequality

$$\begin{aligned} [(\tilde{z}_n, y) - \psi_n]^2 &\leqslant \sum_{i=1}^{\kappa_1} [(U_i, y) - \Phi_i]^2 + [(U_n, y) - \Phi_n]^2 \quad (n = \kappa_1 + 1, \ldots, \kappa_1 + \kappa_2). \\ \sum_{n=\kappa_1 + 1}^{\kappa_1 + \kappa_2} [(\tilde{z}_n, y) - \psi_n]^2 &\leqslant \kappa_2 \sum_{n=1}^{\kappa_1} [(U_i, y) - \Phi_i]^2 + \sum_{n=\kappa_1 + 1}^{\kappa_1 + \kappa_2} [(U_n, y) - \Phi_n]^2. \end{aligned}$$ (5·15)

Since $\{z_n\}$ are linear functions of $\tilde{Y}_{\kappa_1 + 1}, \ldots, \tilde{Y}_{\kappa_1 + \kappa_2}, \psi_n = E(\tilde{z}_n, y) \ (n = \kappa_1 + 1, \ldots)$ are linear functions of parameters $\pi_{\kappa_1 + 1}, \ldots, \pi_{\kappa_1 + \kappa_2}$, a confidence region for these parameters is given by

$$\sum_{n=\kappa_1 + 1}^{\kappa_1 + \kappa_2} [(\tilde{z}_n, y) - \psi_n]^2 \leqslant \kappa_2 \sum_{n=1}^{\kappa_1} [(U_i, y) - \Phi_i]^2 + \sum_{n=\kappa_1 + 1}^{\kappa_1 + \kappa_2} [(U_n, y) - \Phi_n]^2 \leqslant C_2,$$ (5·16)

while a confidence region for the parameters $\pi_1, \ldots, \pi_{\kappa_1}$ is given by

$$\sum_{n=1}^{\kappa_1} [(U_n, y) - \Phi_i]^2 \leqslant C_1.$$ (5·17)

When there are more than two hypotheses in the set M_l, e.g. if there are three hypotheses H_1, H_2, H_3 with the groups of parameters $\pi_1, \ldots, \pi_{\kappa_1}; \pi_{\kappa_1 + 1}, \ldots, \pi_{\kappa_1 + \kappa_2}; \pi_{\kappa_1 + \kappa_2 + 1}, \ldots, \pi_{\kappa_1 + \kappa_2 + \kappa_3}$ with best linear estimates $Y_1, \ldots, Y_{\kappa_1}; Y_{\kappa_1 + 1}, \ldots, Y_{\kappa_1 + \kappa_2}; Y_{\kappa_1 + \kappa_2 + 1}, \ldots, Y_{\kappa_1 + \kappa_2 + \kappa_3}$, we use the same

method as before for the hypothesis H_3 and by combining the hypotheses H_1 and H_2. We then have the following confidence regions for the three groups of parameters:

$$(\pi_1, ..., \pi_{\kappa_1}): \quad \sum_{i=1}^{\kappa_1} [(U_i, y) - \Phi_i]^2 \leqslant C_1,$$

$$(\pi_{\kappa_1+1}, ..., \pi_{\kappa_1+\kappa_2}): \quad \sum_{n=\kappa_1+1}^{\kappa_1+\kappa_2} [(z_n, y) - \psi_n]^2 \leqslant \kappa_2 \sum_{i=1}^{\kappa_1} [(U_i, y) - \Phi_i]^2 + \sum_{i=\kappa_1+1}^{\kappa_1+\kappa_2} [(U_i, y) - \Phi_i]^2 \leqslant C_2,$$

$$(\pi_{\kappa_1+\kappa_2+1}, ..., \pi_{\kappa_1+\kappa_2+\kappa_3}): \quad \sum_{n=\kappa_1+\kappa_2+1}^{\kappa_1+\kappa_2+\kappa_3} [(z_n, y) - \psi_n]^2 \leqslant \kappa_3 \sum_{i=1}^{\kappa_1+\kappa_2} [(U_i, y) - \Phi_i]^2 + \sum_{i=\kappa_1+\kappa_2+1}^{\kappa_1+\kappa_2+\kappa_3} [(U_i, y) - \Phi_i]^2 \leqslant C_3.$$

$$(5 \cdot 18)$$

For the test of the hypotheses

$$
\begin{aligned}
H_1: &\quad \pi_1 = 0, \quad ..., \quad \pi_{\kappa_1} = 0 \\
H_2: &\quad \pi_{\kappa_1+1} = 0, \quad ..., \quad \pi_{\kappa_1+\kappa_2} = 0 \\
H_3: &\quad \pi_{\kappa_1+\kappa_2+1} = 0, \quad ..., \quad \pi_{\kappa_1+\kappa_2+\kappa_3} = 0
\end{aligned}
$$

we consider

$$
\begin{aligned}
\chi_1^2 &= \sum_{i=1}^{\kappa_1} (U_i, y)^2, \quad F_1 = \chi_1^2/(\kappa_1 S_e^2), \\
\chi_2^2 &= \sum_{i=\kappa_1+1}^{\kappa_1+\kappa_2} (U_i, y)^2, \quad F_2 = \chi_2^2/(\kappa_2 S_e^2). \\
\chi_3^2 &= \sum_{i=\kappa_1+\kappa_2+1}^{\kappa_1+\kappa_2+\kappa_3} (U_i, y)^2, \quad F_3 = \chi_3^2/(\kappa_3 S_e^2),
\end{aligned}
$$

$$(5 \cdot 19)$$

Since $\Phi_i = 0$ on the given hypotheses, for all $i = 1, 2, ..., \kappa_1+\kappa_2+\kappa_3$. If the $\alpha \%$ point of F_1, F_2, F_3 (corresponding to G_1, G_2, G_3) are $\epsilon_1, \epsilon_2, \epsilon_3$, then we consider the significance limits of $\frac{1}{\kappa_1} \chi_1^2, \frac{1}{\kappa_2} \chi_2^2, \frac{1}{\kappa_3} \chi_3^2$ for the test of hypotheses H_1, H_2, H_3 respectively as $n_e S_e^2 \epsilon_1$, $n_e S_e^2(\kappa_1 \epsilon_1 + \epsilon_2)$ and $n_e S_e^2(\kappa_1 \epsilon_1 + \kappa_2 \epsilon_2 + \epsilon_3)$. The upper bound of the significance level is then α.

Method (c)

Let $U_1, ..., U_{\kappa_1}; U_{\kappa_1+1}, ..., U_{\kappa_1+\kappa_2}$ be an orthonormal basis of the space formed by the vectors $Y_1, ..., Y_{\kappa_1}; Y_{\kappa_1+1}, ..., Y_{\kappa_1+\kappa_2}$, so that $U_1, ..., U_{\kappa_1}$ is a basis of the space of the vectors $Y_1, ..., Y_{\kappa_1}$. We shall show that when $\kappa_1 < \kappa_2$, it is possible to find orthogonal sets of κ_1 and $\kappa_2 - \kappa_1$ linear functions for the hypothesis H_1 and H_2, so that the inequality $(5 \cdot 3)$ could be used.

In $(5 \cdot 6)$, since the rank of β is at most κ_1, by a transformation of the vectors $\tilde{Y}_1, ..., \tilde{Y}_{\kappa_1}; \tilde{Y}_{\kappa_1+1}, ..., \tilde{Y}_{\kappa_1+\kappa_2}$, the matrix β can be reduced to the triangular form with $\kappa_2 - \kappa_1$ rows of zeros. i.e. the transformed vectors $Y_1', ..., Y_{\kappa_1}'; Y_{\kappa_1+1}', ..., Y_{\kappa_1+\kappa_2}'$ would be given by

$$
\begin{aligned}
Y_{\kappa_1+1}' &= \beta_{1,1}'' U_1 + \beta_{1,2}'' U_2 + ... + \beta_{1,\kappa_1}'' U_{\kappa_1} + \gamma_{1,1}'' U_{\kappa_1+1} + ... + \gamma_{1,\kappa_2}'' U_{\kappa_1+\kappa_2}, \\
Y_{\kappa_1+2}' &= \qquad \beta_{2,2}'' U_2 + ... + \beta_{2,\kappa_1}'' U_{\kappa_1} + \gamma_{2,1}'' U_{\kappa_1+1} + ... + \gamma_{2,\kappa_2}'' U_{\kappa_1+\kappa_2}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
Y_{2\kappa_1+1}' &= \qquad\qquad\qquad\qquad\qquad \gamma_{\kappa_1+1,1}'' U_{\kappa_1+1} + ... + \gamma_{\kappa_1+1,\kappa_2}'' U_{\kappa_1+\kappa_2} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
Y_{\kappa_1+\kappa_2}' &= \qquad\qquad\qquad\qquad\qquad \gamma_{\kappa_2,1}'' U_{\kappa_1+1} + ... + \gamma_{\kappa_2,\kappa_2}'' U_{\kappa_1+\kappa_2}.
\end{aligned}
$$

$$(5 \cdot 20)$$

Let $W_1, ..., W_{\kappa_2-\kappa_1}$ form an orthonormal basis of the vector space of $Y_{2\kappa_1+1}', ..., Y_{\kappa_1+\kappa_2}'$. We then have

$$E\{(W_i, y)\} = \sum_{j=i}^{\kappa_2} \xi_{i,j} \pi_{\kappa_1+j} = \Phi_{\kappa_1+i} \quad (i = 1, ..., \kappa_2 - \kappa_1). \tag{5·21}$$

The hypothesis $\qquad H_2: \quad \pi_{\kappa_1+1} = 0, \quad ..., \quad \pi_{\kappa_1+\kappa_2} = 0,$

implies $\qquad\qquad H_2': \quad \Phi_{\kappa_1+1} = 0, \quad ..., \quad \Phi_{\kappa_1+\kappa_2} = 0,$

since $\Phi_{\kappa_1+1}, ..., \Phi_{\kappa_1+\kappa_2}$ are linear functions of $\pi_{\kappa_1+1}, ..., \pi_{\kappa_1+\kappa_2}$ and have the rank $\kappa_2-\kappa_1$. Now H_2' (i.e. rejection of H_2') implies H_2 (i.e. rejection of H_2). We therefore test for the hypothesis H_2' and reject H_2 only when H_2' is rejected. A confidence region for the parameters $\pi_{\kappa_1+1}, ...,$ $\pi_{\kappa_1+\kappa_2}$ is obtained from $\sum\limits_{i=1}^{\kappa_2-\kappa_1} [(W_i, y) - \Phi_{\kappa_1+i}]^2$ which is distributed independently of

$$\sum_{i=1}^{\kappa_1} [(U_i, y) - \Phi_i]^2,$$

and these sums of squares may be used in equations (5·3) and (4·7).

6. Relative merits of methods (a), (b) and (c)

The method (b) obviously gives a much closer inequality than (a), although the latter has the additional advantage that it gives confidence intervals for all contrasts of the parameters in the hypotheses H_1 and H_2, in the set M_l. However, in most practical cases, the parameters entering into the hypotheses H_1 and H_2 relate to quite distinct characters, and thus there would not, usually, be any need to consider contrasts involving both sets of parameters.

In method (c) we use orthogonal sets of linear functions $Y_1, ..., Y_{\kappa_1}; W_1, ..., W_{\kappa_2-\kappa_1}$, and form χ^2's with κ_1 and $\kappa_2-\kappa_1$ d.f. from these instead of using χ^2's with κ_1 and κ_2 d.f. from the linear functions $Y_1, ..., Y_{\kappa_1}; Y_{\kappa_1+1}, ..., Y_{\kappa_1+\kappa_2}$, which would have given quasi-independent tests for these hypotheses. The loss of degrees of freedom is serious when $\kappa_2-\kappa_1$ is small, as the F-table shows. Thus the method (c) gives a good result only when $\kappa_2-\kappa_1$ is not too small.

On the other hand, the method (b) gives a wide inequality unless κ_1 is small. It is not possible, however, to make a simple quantitative statement about the relative advantages of the methods (b) and (c) without detailed numerical calculations. One could, however, follow a tentative rule for the use of these methods:

$$\kappa_1 \text{ small } \kappa_2-\kappa_1 \text{ small } \qquad \text{use } (b),$$
$$\kappa_1 \text{ small } \kappa_2-\kappa_1 \text{ not small } \qquad \text{use } (b) \text{ or } (c),$$
$$\kappa_1 \text{ not small } \kappa_2-\kappa_1 \text{ small } \qquad \text{use } (b) \text{ or } (c),$$
$$\kappa_1 \text{ not small } \kappa_2-\kappa_1 \text{ not small } \qquad \text{use } (c).$$

REFERENCES

Kimball, A. W. (1951). On dependent tests of significance in the analysis of variance. *Ann. Math. Statist.* **22**, 600.

Nandi, H. K. (1951). On the analysis of variance test. *Bull. Calcutta Statist. Ass.* **3**, 103.

Roy, S. N. & Bose, R. C. (1953). Simultaneous confidence interval estimation. *Ann. Math. Statist.* **24**, 513.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, **40**, 87.

Tukey, J. (1952). Allowances for various types of error rates. Unpublished invited address, Blacksburg meeting of the Institute of Mathematical Statistics, March 1952.

RANK ANALYSIS OF INCOMPLETE BLOCK DESIGNS

III. SOME LARGE-SAMPLE RESULTS ON ESTIMATION AND POWER FOR A METHOD OF PAIRED COMPARISONS*

By RALPH ALLAN BRADLEY

Virginia Agricultural Experiment Station of the Virginia Polytechnic Institute

I. INTRODUCTION

1·1. *Rank analysis of incomplete block designs*

A new method for paired comparisons was discussed in two recent papers (Bradley & Terry, 1952; Terry, Bradley & Davis, 1952). A mathematical model was postulated and tests of significance were developed. The procedures were considered as special cases of a rank analysis of incomplete block designs, and many of the concepts may be extended from the limited considerations of paired comparisons to ranking in incomplete block designs with two or more treatments in an incomplete block. The appropriateness of the model for paired comparisons has been discussed by Hopkins (1954) and also by Bradley (1954a). A large section of tables for a test of significance on the equality of treatment effects was included in the first reference cited and additional tables were prepared by the present author (1954b). For all of the tests considered in the above-referenced work, large-sample approximations to the sampling distributions of the statistics, under the conditions of the null hypotheses, have been given.

1·2. *Review of the method of paired comparisons*

It is necessary to summarize the mathematical model of the method of paired comparisons in order to outline the objectives of this paper.

We postulated true treatment ratings or parameters, $\pi_1, \ldots, \pi_t$, for t treatments in an experiment involving paired comparisons. It was assumed that every $\pi_i \geq 0$, and, for convenience, that† $\sum_i \pi_i = 1$. Further definition followed with the assumption that, when treatment i appears with treatment j in a block, the probability that treatment i obtains the higher rating (or a rank of 1) is $\pi_i/(\pi_i + \pi_j)$. Assuming independence in probability of treatment comparisons, we wrote the likelihood function in its general form as

$$L = \prod_i \pi_i^{a_i} \prod_{i<j} (\pi_i + \pi_j)^{-n}, \tag{1}$$

where

$$a_i = 2n(t-1) - \sum_j' \sum_{k=1}^n r_{ijk}, \tag{2}$$

r_{ijk} is the rank of treatment i in the comparison with treatment j in the kth of n repetitions of the paired comparisons design.

---

† $\sum_i$ and $\prod_i$ will indicate respectively sums and products with $i = 1, \ldots, t$. $\sum_i'$ will mean that one value of i that appears in the argument of the summation is omitted. $\prod_{i<j}$ or $\prod_{i \neq j}$ represent products. $i = 1, \ldots, t, j = 1, \ldots, t, i < j$ or $i \neq j$ respectively. Departures from these conventions will be specified.

The method of maximum likelihood was used to obtain estimators $p_1 \ldots p_t$ of the parameters, $\pi_1, \ldots, \pi_t$, and likelihood ratio tests have been used throughout discussions based on the model set down. A general class of tests of the null hypothesis.

$$H_0: \quad \pi_i = 1/t \quad (i = 1, \ldots, t),$$

against the alternative hypothesis,

$$H_a: \quad \pi_i = \pi(h) \quad (h = 1, \ldots, m; \ i = s_{h-1}+1, \ldots, s_h),$$

where $s_0 = 0$, $s_m = t$ and $\sum_{h=1}^{m}(s_h - s_{h-1})\pi(h) = 1$, was formulated. These are, of course, tests of the indistinguishability of treatment effects on some attribute, perhaps colour or flavour, of the treatments. Two special cases were considered by Bradley & Terry (1952) and the specialization comes under H_a.

Case (i): The hypothesis H_a becomes

$$H_1: \quad \text{No } \pi_i \text{ is assumed equal to any } \pi_j \ (i \neq j);$$

that is, in H_a, $m = t$.

The normal equations resulting from the use of the method of maximum likelihood reduced to

$$\frac{a_i}{p_i} - n \sum_j{}' (p_i + p_j)^{-1} = 0 \quad (i = 1, \ldots, t) \tag{3}$$

and

$$\sum_i p_i = 1. \tag{4}$$

In this paper we shall be particularly interested in the statistic* $T = -2 \ln \lambda_1$, where λ_1 is the likelihood ratio for this special test. In this case

$$T = nt(t-1)\ln 2 - 2B_1 \ln 10, \tag{5}$$

with†

$$B_1 = n \sum_{i<j} \log(p_i + p_j) - \sum_i a_i \log p_i. \tag{6}$$

Case (ii): H_a becomes

$$H_2: \quad \pi_i = \pi \ (i = 1, \ldots, s); \qquad \pi_i = (1 - s\pi)/(t - s) \ (i = s+1, \ldots, t).$$

The general hypothesis H_a is thus restricted to the case in which $m = 2$. The estimator p of π was given as

$$p = \frac{ns(4t - s - 3) - 2\sum_{i=1}^{s}\sum_j{}'\sum_{k=1}^{n} r_{ijk}}{ns(5st - 2t^2 - 6s + 3t) - 2(2s - t)\sum_{i=1}^{s}\sum_j{}'\sum_{k=1}^{n} r_{ijk}}. \tag{7}$$

A test statistic and an approximate test procedure were set forth in the reference. It will not be necessary to review these methods in view of the remarks that follow in §II.

Abelson & Bradley (1954) considered factorial arrangements of treatments imposed on the paired comparisons design. We shall not consider that situation in this paper.

1·3. *Objectives*

The objectives of this paper evolve from the need of considering the behaviour of the developed tests of significance and estimates of population parameters when the assumptions of the null hypotheses may not be true. We shall be interested in the power of the test

* log and ln indicate common and natural logarithms respectively.
† B_1 is the statistic tabulated for small samples. See Bradley & Terry (1952) and Bradley (1954b).

based on T in (5), in the reliability of the estimators p_i of π_i defined by (2) and (3), and in comparisons of the power of this test procedure with those of other possible procedures.

In the initial reference, it has been shown that T, in (5), under H_0, has the distribution of χ^2 with $(t-1)$ degrees of freedom for large samples. The difficulty to be expected in attempting an exact evaluation of the power of the test based on T for small samples, even for very restricted sets of alternative values of the parameters, was also discussed. Accordingly, we limit our objectives here to a large-sample evaluation and comparison of power functions and to the estimation of variances and covariances using large-sample results.

We shall show that Case (ii) yields the 'sign test', the properties of which are well known, and we may then limit our attention to Case (i).

II. Case (ii) and the sign test

2·1. *Case (ii)*

We refer to H_2: $\pi_i = \pi$ $(i = 1, ..., s)$; $\pi_i = (1 - s\pi), (t - s)$ $(i = s + 1, ..., t)$ and the estimator p of π given in (7). It was not originally noted that the test procedure here reverts back to a binomial test.

Comparisons of treatments in the first group of s treatments yield contributions to the sums of ranks in (7) of 3. Then, if X is defined to be the number of times a treatment of the first group ranks above (obtains rank 1) a treatment of the second group,

$$\sum_{i-1}^{s} \sum_{j}' \sum_{k=1}^{n} r_{ijk} = \frac{3ns(s-1)}{2} + 2ns(t-s) - X. \tag{8}$$

Substitution of (8) in (7) yields

$$p = X/[(2s-t)X + ns(t-s)^2]. \tag{9}$$

Now, from the model for paired comparisons, the probability that a treatment i of the first group ranks above a treatment j of the second group in any of the n repetitions is

$$P(r_{ijk}=1) = \frac{\pi}{\pi + \dfrac{1-s\pi}{t-s}} = \frac{\pi(t-s)}{1 + (t-2s)\pi}, \tag{10}$$

$$(i = 1, ..., s; \; j = s+1, ..., t; \; k = 1, ..., n).$$

When we substitute the estimator p of π given in (9) in the form (10), we obtain the estimated probability,

$$\text{Est } P(r_{ijk}=1) = X/[ns(t-s)]. \tag{11}$$

There are $ns(t-s)$ comparisons of treatments of the first group with treatments of the second group, and it is now apparent that this special case reduces to a consideration of $ns(t-s)$ binomial trials. The test procedures reduce to those of the binomial or sign test.

2·2. *The sign test*

The properties of the sign test have been thoroughly investigated. Hemelrijk (1952) notes that it is probably the oldest test in existence and refers to an application by Arbuthnot in 1710. The variance and approximate normality of X are discussed in elementary texts, and exact power function evaluations can be obtained for small samples from tables of the binomial distribution or of the incomplete beta function. The power of the sign test for large samples is based on the approximate normality of the statistic.

Dixon & Mood (1946) showed that the relative efficiency of the sign test in comparison with the t test under assumptions suitable for the valid application of the latter test is $2/\pi$. Later, Dixon (1953) prepared small-sample tables of power efficiencies that indicated that the sign test compares more favourably with the t-test than indicated by the asymptotic value, $2/\pi$.

In considering applications of Case (ii), it is preferable to go directly to the use of the sign test based on the $ns(t-s)$ comparisons of treatments in the first group with those of the second group. It will usually be sufficient to estimate $P(r_{i,l} = 1)$ and not necessary to consider estimating π itself. If required, the variance of p, the estimator of π, may be obtained from the variance of X in an approximation through the use of usual formulae for the variance of a ratio.

We shall devote the remainder of this paper to a consideration of Case (i).

III. ESTIMATION

3·1. *Asymptotic distribution of $(t-1)$ maximum-likelihood estimators*

We shall require the large-sample distribution of the maximum-likelihood estimators, $p_1, \ldots, p_t$, of Case (i) and their asymptotic variances and covariances. These results will be of interest in themselves and useful in the development of subsequent sections of this paper.

Some new notation will assist in this discussion. Let

$$f(x, \pi) = \prod_i \pi_i^{x_i} \prod_{i<j} (\pi_i + \pi_j)^{-1}, \tag{12}$$

where we use x and π to represent vectors, $(x_1, \ldots, x_t)$ and $(\pi_1, \ldots, \pi_t)$. Now x_i is the number of times treatment i obtains a ranking of unity in a repetition of paired comparisons. If $x_{i(k)}$ is the observation on x_i in the kth of n repetitions, the association with a_i of (1) is

$$\sum_{k=1}^{n} x_{i(k)} = a_i \quad (i = 1, \ldots, t). \tag{13}$$

The likelihood function L in (1) is now $\prod_{k=1}^{n} f(x_{(k)}, \pi)$. It is also convenient to define

$$\left. \begin{array}{l} \lambda_{ii} = \dfrac{1}{\pi_i} {\sum_h}' \pi_h(\pi_i + \pi_h)^{-2} \quad (i = 1, \ldots, t), \\[2mm] \lambda_{ij} = -(\pi_i + \pi_j)^{-2} \quad (i \neq j;\, i, j = 1, \ldots, t). \end{array} \right\} \tag{14}$$

We shall require the means, variances and covariances of $x_1, \ldots, x_t$. Let x_{ij} be an indicator variate with the value unity if treatment i ranks above treatment j and zero otherwise. Then $x_i = \sum_j x_{ij}$. x_{ij} is a binomial variate with expectation $\pi_i(\pi_i + \pi_j)^{-1}$, and variance $\pi_i \pi_j(\pi_i + \pi_j)^{-2}$. The variates x_{ij} making up the sum x_i are independent in probability and it follows that

$$E(x_i) = \pi_i {\sum_h}' (\pi_i + \pi_h)^{-1} \quad (i = 1, \ldots, t) \tag{15}$$

$$V(x_i) = \pi_i {\sum_h}' \pi_h(\pi_i + \pi_h)^{-2} = \pi_i^2 \lambda_{ii} \quad (i = 1, \ldots, t), \tag{16}$$

and

$$\text{cov}\,(x_i, x_j) = -\pi_i \pi_j(\pi_i + \pi_j)^{-2} = \pi_i \pi_j \lambda_{ij} \quad (i \neq j;\, i, j = 1, \ldots, t). \tag{17}$$

The parameters $\pi_1, \ldots, \pi_t$ are not independent but subject to the restriction $\sum_i \pi_i = 1$. Accordingly we may regard $p_1, \ldots, p_{t-1}$ as maximum-likelihood estimators of the independent parameters, $\pi_1, \ldots, \pi_{t-1}$, taking $\pi_t = 1 - \sum_{i=1}^{t-1} \pi_i$. Then $\sqrt{n}\,(p_1 - \pi_1), \ldots, \sqrt{n}\,(p_{t-1} - \pi_{t-1})$ have a joint limiting normal distribution with zero means subject to the verification of certain regularity conditions. The required regularity conditions are quite well known and given, for example, by Cramér (1946, §33·3, p. 500) and Chanda (1954), who fully states the conditions but for the continuous case. We have verified the conditions for the likelihood function L of (1) expressed in terms of $f(x, \pi)$ of (12), but these demonstrations are omitted for brevity.

The dispersion matrix of the joint limiting normal distribution of our $(t-1)$ estimators is the matrix $[\lambda'_{ij}]^{-1}$ with the definition of λ'_{ij} below. We need only note that

$$\frac{\partial \ln f}{\partial \pi_i} = \frac{x_i}{\pi_i} - \frac{x_t}{\pi_t} - \sum_h{}' (\pi_i + \pi_h)^{-1} + \sum_h{}' (\pi_t + \pi_h)^{-1} \quad (i = 1, \ldots, (t-1)). \tag{18}$$

and, from (15), that

$$E\left[\left(\frac{\partial \ln f}{\partial \pi_i}\right)\left(\frac{\partial \ln f}{\partial \pi_j}\right)\right] = \mathrm{cov}\left[\left(\frac{x_i}{\pi_i} - \frac{x_t}{\pi_t}\right)\left(\frac{x_j}{\pi_j} - \frac{x_t}{\pi_t}\right)\right] = \lambda'_{ij} \quad (i, j = 1, \ldots, (t-1)),$$

thereby defining λ'_{ij}. It follows from (16) and (17) that

$$\lambda'_{ij} = \lambda_{ij} - \lambda_{it} - \lambda_{jt} + \lambda_{tt}. \tag{19}$$

The matrix $[\lambda'_{ij}]$ is non-negative definite since it is a dispersion matrix and it is positive definite since $x_1, \ldots, x_{t-1}$, and hence $\dfrac{\partial \ln f}{\partial \pi_1}, \ldots, \dfrac{\partial \ln f}{\partial \pi_{t-1}}$, are free of linear restrictions.

The conclusion, with the established validity of the necessary conditions, is that

$\sqrt{n}\,(p_1 - \pi_1), \ldots, \sqrt{n}\,(p_{t-1} - \pi_{t-1})$ *have for large samples the multivariate normal distribution with zero means and dispersion matrix* $[\lambda'_{ij}]^{-1}$.

3·2. *Variances and covariances of estimators*

In the preceding section, the parameter π_t was considered to be a function of the remaining $(t-1)$ parameters. That process introduced asymmetry into the elements of the dispersion matrix, $[\lambda'_{ij}]^{-1}$. This lack of symmetry is essentially artificial and will now be removed.

The $t \times t$ matrix, $[\lambda_{ij}]$, is singular in view of the definitions (14) and the fact that

$$\pi_i \lambda_{ii} + \sum_j{}' \pi_j \lambda_{ij} = 0. \tag{20}$$

Then, if the elements of the last row and then of the last column of the matrix $[\lambda_{ij}]$ are subtracted from corresponding elements of the remaining rows and columns respectively,

$$|\lambda_{ij}| = \begin{vmatrix} [\lambda'_{ij}] & [\lambda_{it} - \lambda_{tt}]' \\ [\lambda_{jt} - \lambda_{tt}] & \lambda_{tt} \end{vmatrix} = 0,$$

where $[\lambda_{it} - \lambda_{tt}]'$ and $[\lambda_{jt} - \lambda_{tt}]$ are respectively column and row vectors of $(t-1)$ elements. It follows that

$$|\lambda'_{ij}| = \begin{vmatrix} [\lambda'_{ij}] & [\lambda_{it} - \lambda_{tt}]' \\ [\lambda_{jt} - \lambda_{tt}] & 1 + \lambda_{tt} \end{vmatrix}. \tag{21}$$

Now, by reversing the process and adding the elements of the last row and then the last column to corresponding elements of the remaining rows and columns in (21), we have

$$|\lambda'_{ij}| = |\lambda_{ij} + 1| = \begin{vmatrix} [\lambda_{ij} + 1] & [1]' \\ [0] & 1 \end{vmatrix} = -\begin{vmatrix} [\lambda_{ij}] & [1]' \\ [1] & 0 \end{vmatrix}, \tag{22}$$

where [1] and [1]' are respectively row and column vectors of t unit elements. The last step again depends on the result that $|\lambda_{ij}| = 0$. In the same way it can be shown that the cofactor of λ_{ij} in the extreme right-hand side of (22) is equal to the cofactor of λ'_{ij} in $|\lambda'_{ij}|$. Then, if σ_{ij} is the covariance of $\sqrt{n}(p_i - \pi_i)$ and $\sqrt{n}(p_j - \pi_j)$ $(i, j = 1, ..., (t-1))$.

$$\sigma_{ij} = \frac{\text{cofactor of } \lambda_{ij} \text{ in } \begin{vmatrix} [\lambda_{ij}] & [1]' \\ [1] & 0 \end{vmatrix}}{\begin{vmatrix} [\lambda_{ij}] & [1]' \\ [1] & 0 \end{vmatrix}}. \tag{23}$$

But the formulation of the model for paired comparisons is symmetric in the parameters $\pi_1, ..., \pi_t$ and the estimators $p_1, ..., p_t$, and (23) applies for all variances and covariances with $i, j = 1, ..., t$ on the basis of its symmetry.

The variances and covariances obtainable from (23) are simply the elements of the t-square principal minor of the inverse of the matrix of the determinant in the denominator of (23). For small values of t, this inverse matrix may be evaluated by elementary methods when $\pi_1, ..., \pi_t$ are specified or replaced by estimates. However, since $|\lambda_{ij}| = 0$, the usual Doolittle methods of matrix inversion break down. Then for larger values of t, it appears desirable to invert $[\lambda'_{ij}]$ specified through (19) using a Doolittle method, thus obtaining the required variances and covariances with $i, j = 1, ..., (t-1)$. The remaining variance and covariances associated with $\sqrt{n}(p_t - \pi_t)$ are determined through the relationship

$$\sqrt{n}(p_t - \pi_t) = -\sum_{i=1}^{t-1} \sqrt{n}(p_i - \pi_i),$$

so that

$$\sigma_{tt} = \sum_{i=1}^{t-1}\sum_{j=1}^{t-1} \sigma_{ij} \quad \text{and} \quad \sigma_{it} = -\sum_{j=1}^{t-1} \sigma_{ij} \quad (i = 1, ..., (t-1)).$$

Since it has been established in §3·1 that $\sqrt{n}(p_1 - \pi_1), ..., \sqrt{n}(p_{t-1} - \pi_{t-1})$ have asymptotically a multivariate normal distribution, and since $\sqrt{n}(p_t - \pi_t)$ is a linear function of those variates, we may now state that

$\sqrt{n}(p_1 - \pi_1), ..., \sqrt{n}(p_t - \pi_t)$ *have, for large values of* n, *the singular multivariate normal distribution of* $(t-1)$ *dimensions in a space of* t *dimensions with zero means and dispersion matrix,* $[\sigma_{ij}]$, *defined through* (23).

In general, for large samples, we may take $p_1, ..., p_t$ to be jointly normally distributed with means $\pi_1, ..., \pi_t$, and dispersion matrix $[\sigma_{ij}]/n$. Then any linear function $\sum_i b_i p_i$ may be taken to be normal for large samples with mean $\sum_i b_i \pi_i$ and variance $\sum_{i<j} b_i b_j \sigma_{ij}/n$. In particular, we may be interested in orthogonal linear comparisons of the sort often considered in the analysis of variance.

Estimated variances and covariances will usually be required. Through the consistency of maximum-likelihood estimators, we can define $\hat{\lambda}_{ij}$ $(i, j = 1, ..., t)$ to be the same functions of $p_1, ..., p_t$ as λ_{ij} are of $\pi_1, ..., \pi_t$ as defined in (14). Then $[\hat{\sigma}_{ij}]$, the sample dispersion matrix, is the same function of $\hat{\lambda}_{ij}$ as $[\sigma_{ij}]$ is of λ_{ij} in definition (23).

It has been shown (Bradley, 1953) that the functions of the parameters. $\ln \pi_1, ..., \ln \pi_t$, determine in a sense location points for the t treatments on an arithmetic scale. Consequently, there may be some interest in obtaining

$$\sigma_{ij}(\ln p_i, \ln p_j) \approx \frac{\sigma_{ij}}{n\pi_i \pi_j} \quad (i, j = 1, ..., t), \tag{24}$$

as given, for example, by Hald (1952). If common logarithms are used, a factor. $(0.4343)^2$, is required in the right-hand member of (24).

IV. Large-sample distributions of T

4·1. *T expanded in a series*

It has been shown (Bradley & Terry, 1952) that, under the conditions of the null hypothesis of Case (i) of § 1·2, T as defined in (5) has, asymptotically with n, the χ^2-distribution with $(t-1)$ degrees of freedom. In order to investigate the power of the test procedure based on T and to consider the efficiency of the method in comparison with other test procedures, we require the distribution of T under the alternative hypothesis H_a of Case (i). In this section we shall show that T has the same limiting distribution as a certain sum of squares.

Let

$$y_i = t\left(p_i - \frac{1}{t}\right) \quad (i = 1, ..., t), \tag{25}$$

and note that from (2)

$$a_i = \tfrac{1}{2}n(1 + y_i) \sum_j{}' [1 + \tfrac{1}{2}(y_i + y_j)]^{-1} \tag{26}$$

for all i, and

$$\sum_i a_i = \tfrac{1}{2}nt(t-1). \tag{27}$$

Substitution in (5) through the use of (6) and upon reduction based on (26) and (27) yields

$$T = n \sum_i (1 + y_i)[\ln (1 + y_i)] \sum_j{}' [1 + \tfrac{1}{2}(y_i + y_j)]^{-1} - 2n \sum_{i<j} \ln [1 + \tfrac{1}{2}(y_i + y_j)]^{-1}. \tag{28}$$

It is next possible, by expanding the right-hand member of (28) in a power series in the variates, to show that

$$T = \tfrac{1}{4}nt \sum_i y_i^2 + R(y_i), \tag{29}$$

where $R(y_i)$ depends on higher powers of the variates than the second. In turn, it can be shown that

$$| R(y_i) | \leqslant \tfrac{1}{12}n(5t-8) \sum_i | y_i |^3 + \tfrac{1}{12}n(9t-14) \sum_i y_i^4 + \tfrac{1}{6}n(t-2) \sum_i | y_i |^5 + \tfrac{1}{12}n \sum_{i<j} | y_i + y_j |^3$$

$$+ n \sum_i [| y_i | + \tfrac{1}{2}y_i^2 + \tfrac{5}{6} | y_i |^3 + \tfrac{1}{3}y_i^4] \sum_j{}' \frac{(y_i + y_j)^2}{4[1 + \tfrac{1}{2}(y_i + y_j)]}. \tag{30}$$

This result, (30), is obtained through algebraic manipulation involving the use of the general relations

$$| \ln (1 + x) - x + \tfrac{1}{2}x^2 | \leqslant | \tfrac{1}{3}x |^3, \quad (1 + x)^{-1} = 1 - x + x^2/(1 + x),$$

and the particular relations

$$\sum_j{}' y_j = -y_i \quad \text{and} \quad 2 \sum_{i<j} y_i y_j = -\sum_i y_i^2.$$

Let us suppose that $\sqrt{n}\,y_i\,(i=1,...,t)$ have limiting distributions and this will be demonstrated under desired conditions. Assuming this result, we may then state that $n^{\frac{3}{2}}y_i$, for any $\epsilon>0$, converges in probability to zero, for it is an easily proved theorem in probability that, if (X_N) represents a sequence of random variables with a limiting distribution function and if (α_N) is a sequence of constants approaching zero as $N\to\infty$, then $(\alpha_N X_N)$ is a sequence of random variables converging in probability to zero as $N\to\infty$. It follows from Slutsky's theorem (Cramér, 1946, §20·6, p. 255) that, if $\sqrt{n}\,y_i$ have limiting distributions, $R(y_i)$ converges stochastically to zero. This is sufficient (Cramér, 1946, §20·6, p. 254) to state that T has the same limiting distribution as $\frac{1}{4}nt\sum_i y_i^2$.

4·2. Limiting distribution of y_i under H_a

The normal equations for the maximum-likelihood estimators were given in (3) and (4), and these relations are useful in obtaining the limiting distribution of $\sqrt{n}\,y_i$, defined in (25) under modified specifications of $\pi_1,...,\pi_t$. From (3) or (26), we have

$$a_i/n = \frac{1}{2}(1+y_i)\sum_j{}'[1+\frac{1}{2}(y_i+y_j)]^{-1} \quad (i=1,...,t). \tag{31}$$

Expansion in (31), using the negative binomial and the result that $\sum_j{}' y_j = -y_i$, and multiplication by $\sqrt{n}$ lets us write

$$\sqrt{n}\left[\frac{a_i}{n}-\frac{1}{2}(t-1)\right]=\frac{1}{4}t\sqrt{n}\,y_i-\frac{(t-2)}{4\sqrt{n}}(\sqrt{n}\,y_i)^2+\frac{1}{8\sqrt{n}}\sum_j{}'\frac{(1+y_i)(\sqrt{n}\,y_i+\sqrt{n}\,y_j)^2}{[1+\frac{1}{2}(y_i+y_j)]}. \tag{32}$$

Now, from an argument similar to that used in §4·1, if $\sqrt{n}\,y_i$ has a limiting distribution, it has the same limiting distribution as

$$\frac{4\sqrt{n}}{t}\left[\frac{a_i}{n}-\frac{1}{2}(t-1)\right]. \tag{33}$$

Let us now redefine the parameters so that

$$\pi_i=\frac{1}{t}+\frac{\delta_{in}}{\sqrt{n}} \quad (i=1,...,t), \tag{34}$$

where δ_{in} represents a sequence of constants converging to δ_i as $n\to\infty$. This redefinition of the parameters is an artifice that permits us to find the limiting distribution of T under H_a and thence the power of the test procedure. This essentially means that we are investigating the asymptotic power for alternatives in the locality of the parameter point determined by the null hypothesis. This is necessary in order that we do not merely show that the test procedure is consistent. Under these conditions we require the limiting distribution of the variate defined in (33).

Let $\bar{x}_{ij}$ be the average over the n repetitions of the binomial variate, x_{ij}, defined in §3·1. Then, through the definitions of x_i and $x_{i(k)}$ and (13), we have

$$a_i=n\sum_j{}'\bar{x}_{ij}, \tag{35}$$

and the variates, $\bar{x}_{i1},...,\bar{x}_{it}$, omitting $\bar{x}_{ii}$, are independent.

Since $E(x_{ij})=\pi_i(\pi_i+\pi_j)^{-1}$, we may define $\phi_n(\tau)$ to be the characteristic function of $2\sqrt{n}\,(\bar{x}_{ij}-\frac{1}{2})$, and, following the method of Cramér (1946) in §16·4 of his book, we have

$$\phi_n(\tau)=\left[\frac{\pi_j\exp\left\{-\sqrt{-1}\dfrac{\tau}{\sqrt{n}}\right\}}{(\pi_i+\pi_j)}+\frac{\pi_i\exp\left\{\sqrt{-1}\dfrac{\tau}{\sqrt{n}}\right\}}{(\pi_i+\pi_j)}\right]^n. \tag{36}$$

Expansion of the exponentials in (36) yields

$$\phi_n(\tau) = \left[1 + \left(\frac{\pi_i - \pi_j}{\pi_i - \pi_j} \right)_\checkmark - 1 \frac{\tau}{n} - \frac{\tau^2}{2n} + \left(\frac{\pi_i \gamma^3 - \pi_i \theta^3}{\pi_i + \pi_j} \right) \frac{\tau^3}{6n^{\frac{3}{2}}} \right]^n, \tag{37}$$

where γ and θ are real or complex quantities less than unity in modulus.

We now substitute from (34) for the parameters in (37) and write

$$\phi_n(\tau) = \left[1 + \checkmark - 1 \frac{t\tau}{2n} (\delta_i - \delta_j) + \frac{H}{n} - \frac{\tau^2}{2n} (1 + W) \right]^n,$$

where $H = \checkmark - 1 \frac{t\tau}{2} (\delta_{in} - \delta_i - \delta_{jn} + \delta_j) - \checkmark - 1 \frac{t^2\tau}{4\sqrt{n}} (\delta_{in}^2 - \delta_{jn}^2) \left[1 + \frac{t}{2\sqrt{n}} (\delta_{in} + \delta_{jn}) \right]^{-1}.$

and $W = \frac{\tau}{3\sqrt{n}} \left[\frac{1}{t} (\theta^3 - \gamma^3) + \frac{\delta_{in} \theta^3}{\sqrt{n}} - \frac{\delta_{jn} \gamma^3}{\sqrt{n}} \right] \left[\frac{2}{t} + \frac{\delta_{in} + \delta_{jn}}{\sqrt{n}} \right]^1.$

From the convergence of δ_{in} to δ_i and the forms of H and W, for any $\epsilon, \eta > 0$, there exists $n_0(\epsilon, \eta)$ such that, when $n > n_0$, $|H| < \epsilon$ and $|W| < \eta$ for fixed τ. It follows that

$$\exp\left[-\tfrac{1}{2}\tau^2 (1 + \eta) + \tfrac{1}{2}\checkmark - 1 t\tau (\delta_i - \delta_j) - \epsilon \right] \leqslant \lim_{n \to \infty} \phi_n(\tau)$$

$$\leqslant \exp\left[-\tfrac{1}{2}\tau^2 (1 - \eta) + \tfrac{1}{2}\checkmark - 1 t\tau (\delta_i - \delta_j) + \epsilon \right],$$

and hence $\lim_{n \to \infty} \phi_n(\tau) = \exp\left[-\tfrac{1}{2}\tau^2 + \tfrac{1}{2}\checkmark - 1 t\tau (\delta_i - \delta_j) \right].$ (38)

This result (38) is sufficient (see, for example, Cramér, 1946, § 10·4) to state that $2\sqrt{n}(x_{ij} - \tfrac{1}{2})$ has a normal limiting distribution with mean $\tfrac{1}{2}t(\delta_i - \delta_j)$ and unit variance.

We may obtain the variances and covariances of $4\sqrt{n}[a_i/n - \tfrac{1}{2}(t - 1)]/t$ through use of the relation (35). It is easy to show that

$$\frac{4\sqrt{n}}{t} \left[\frac{a_i}{n} - \tfrac{1}{2}(t - 1) \right] = \frac{2}{t} \sum_j{}' [2\sqrt{n}(\bar{x}_{ij} - \tfrac{1}{2})], \tag{39}$$

and we note also that by definition $\bar{x}_{ji} = 1 - \bar{x}_{ij}.$ (40)

The mean of the left-hand member of (39) in the limit is now seen to be $\sum_j{}' (\delta_i - \delta_j) = t\delta_i$, since $\sum_j{}' \delta_j = -\delta_i$. Through the variance of $2\sqrt{n}(\bar{x}_{ij} - \tfrac{1}{2})$ and the independence of the variates in the sum in (39), the limiting variance of the left-hand member of (39) is $4(t - 1)/t^2$. $\bar{x}_{ij}$ and $\bar{x}_{gh}$ are independent unless $i = g$ and $j = h$ or $i = h$ and $j = g$. Then in view of (40), the co-variance of $4\sqrt{n}[a_i/n - \tfrac{1}{2}(t - 1)]/t$ and $4\sqrt{n}[a_j/n - \tfrac{1}{2}(t - 1)]/t$ $(i \neq j)$ is $-4/t^2$. It follows that $4\sqrt{n}[a_1/n - \tfrac{1}{2}(t - 1)]/t, \ldots, 4\sqrt{n}[a_t - \tfrac{1}{2}(t - 1)]/t$, and consequently $\sqrt{n}y_1, \ldots, \sqrt{n}y_t$, have a joint limiting normal distribution with means $t\delta_1, \ldots, t\delta_t$, and equal variances and covariances $4(t - 1)/t^2$ and $-4/t^2$, respectively under the conditions (34) on the parameters as $n \to \infty$. Since $\sum_i y_i = 0$, the limiting distribution is necessarily singular.

It is to be noted that this same limiting distribution can apparently be obtained by using the definition (34) for the parameters and relying on the use of the joint limiting distribution of $\sqrt{n}(p_1 - \pi_1), \ldots, \sqrt{n}(p_t - \pi_t)$ developed in § 3·2 above. If π_i is replaced by $t^{-1} + n^{-\frac{1}{2}}\delta_{in}$ in the definitions of λ_{ij} in (14), and if λ_{ij} is then replaced by $\lim_{n \to \infty} \lambda_{ij}$ in (23), the correct variances and covariances would be obtained for $\sqrt{n}y_1, \ldots, \sqrt{n}y_t$. That this procedure is valid has been proved by Wald (1943) but we have obtained the results directly.

4·3. *Limiting distribution of T under $H_a: \pi_i = t^{-1} + n^{-\frac{1}{2}}\delta_{in}$*

We define
$$u_i = \tfrac{1}{2}\sqrt{t}\sqrt{n}\,y_i, \tag{41}$$

and, from § 4·1, T has the same limiting distribution as $\sum_i u_i^2$; $u_1, ..., u_t$ have a joint limiting distribution that is multivariate normal with means, $\tfrac{1}{2}t^{\frac{1}{2}}\delta_1, ..., \tfrac{1}{2}t^{\frac{1}{2}}\delta_t$, variances $(t-1)/t$ and covariances $-1/t$. Further, $\sum_i u_i = 0$.

The Helmert transformation is used to transform to new variables, $z_1, ..., z_t$, with $z_t = 0$, and hence T has the same limiting distribution as $\sum_{i=1}^{t-1} z_i^2$. It is easily verified that $z_1, ..., z_{t-1}$ are independent, have unit variances, and have limiting normal distributions under H_a with means, say, $\theta_1, ..., \theta_{t-1}$, where $\sum_{i=1}^{t-1} \theta_i^2 = \tfrac{1}{4}t^3 \sum_i \delta_i^2$. This is sufficient to permit the final conclusion that $\sum_{i=1}^{t-1} z_i^2$ and T *have a limiting distribution under* $H_a: \pi_i = t^{-1} + n^{-\frac{1}{2}}\delta_{in}$ *as* $n \to \infty$, *that is, the non-central χ^2-distribution with $(t-1)$ degrees of freedom and parameter of non-centrality,* $\lambda = \tfrac{1}{4}t^3 \sum_i \delta_i^2$. λ is defined if we reiterate that the limiting distribution function of T under H_a is obtained by integrating the non-central χ^2-density,

$$f(T) = \frac{e^{-\frac{1}{2}T}e^{-\frac{1}{2}\lambda}}{2^{\frac{1}{2}(t-1)}} \sum_{h=0}^{\infty} \frac{T^{\frac{1}{2}(t-1)+h-1}\lambda^h}{\Gamma[\frac{1}{2}(t-1)+h]\,2^h h!}. \tag{42}$$

The power of the paired comparisons test based on T is asymptotically given by

$$\beta(\lambda \mid \alpha, t-1, \infty) = \int_{\chi^2_{\alpha, t-1}}^{\infty} f(T)\,dT, \tag{43}$$

where $\chi^2_{\alpha, t-1}$ is the α-level significance value of a central χ^2-distribution with $(t-1)$ degrees of freedom. Fix (1949) provided tables for $\alpha = 0.01$ and $\alpha = 0.05$ with λ tabulated for given values of β and for given degrees of freedom. Pearson & Hartley (1951) provided charts for given degrees of freedom and $\alpha = 0.01$ and 0.05, whereon β is plotted against ϕ. These charts are for the non-central F-distribution but may be used for the non-central χ^2-distribution if the denominator degrees of freedom of F are taken to be infinite. The notation $\beta(\lambda \mid \alpha, \nu_1, \nu_2)$ was used by Pearson & Hartley and has been modified for our special case in (43).

$$\phi = \sqrt{(\lambda/t)} = \tfrac{1}{2}t\sqrt{\sum_i \delta_i^2}. \tag{44}$$

Tang (1938) presented tables based on ϕ that may also be used to evaluate the integral (43).

For large n, an adequate approximation to the power of the paired comparisons test may be obtained by assuming the non-central χ^2-distribution and taking the parameter λ to be

$$\lambda = \tfrac{1}{4}nt^3 \sum_i (\pi_i - 1/t)^2. \tag{45}$$

When H_0 is true, $\pi_i = 1/t$ and $\delta_{in} = \delta_i = 0$ $(i = 1, ..., t)$. Accordingly, $\lambda = 0$ and T has then the central χ^2-distribution with $(t-1)$ degrees of freedom for large n. This simply re-establishes the approximate large-sample distribution for the test of H_0 given by Bradley & Terry (1952).

V. Comparisons of powers of tests

5·1. *Comparison of the test for paired comparisons with a multi-binomial test*

One of the chief uses of an investigation of the power of a statistical test is an assessment of the merits of the test in contrast with alternative procedures. In this and the following section we shall compare the test based on T with two other possible methods.

The model formulated for the method of paired comparisons is not the most general one of its form possible. We could postulate parameters π_{ij}, the probability of treatment i being rated above treatment j in the comparison of these two treatments. Then the n repetitions of the paired comparisons design could be regarded as $\frac{1}{2}t(t-1)$ unrelated binomial tests, each test depending on n trials.

Let us consider the comparison of treatment i with treatment j. The usual test of the null hypothesis, $\pi_{ij} = \frac{1}{2}$, is based on the statistic,

$$S_{ij} = 4n(p_{ij} - \tfrac{1}{2})^2, \tag{46}$$

where p_{ij} is the estimator of π_{ij}. Further,

$$S_{ij} = 2n(p_{ij} - \tfrac{1}{2})^2 + 2n(p_{ji} - \tfrac{1}{2})^2 = u_{ij}^2 + u_{ji}^2$$

and u_{ij} and u_{ji} correspond to variates u_i defined in (41). This binomial test is only a special case, $t = 2$, of the paired comparisons method being discussed in this paper. Then, under the null hypothesis, S_{ij} has a limiting χ^2-distribution with 1 degree of freedom. (This result is of course well known.) Further, if the alternative hypothesis is expressed as

$$\pi_{ij} = \tfrac{1}{2} + \mu_{ijn}/\sqrt{n}$$

with μ_{ijn} converging to μ_{ij} as $n \to \infty$, S_{ij} has a limiting non-central χ^2-distribution with 1 degree of freedom and parameter of non-centrality, $\lambda_{ij} = 2(\mu_{ij}^2 + \mu_{ji}^2)$, $\pi_{ij} = 1 - \pi_{ji}$ and $\mu_{ij} = -\mu_{ji}$. The distribution of S_{ij} under the alternative hypothesis follows as a special case of the theory developed for paired comparisons.

Suppose that the model for paired comparisons is appropriate. Then, $\pi_{ij} = \pi_i/(\pi_i + \pi_j)$ and, if $\pi_i = t^{-1} + n^{-\frac{1}{2}}\delta_{in}$ as before,

$$\mu_{ijn} = \tfrac{1}{4}t(\delta_{in} - \delta_{jn}) - \frac{t^2}{4\sqrt{n}}\delta_{in}(\delta_{in} + \delta_{jn}) + \frac{t^3}{8\sqrt{n}}\left(\frac{1}{t} + \frac{\delta_{in}}{\sqrt{n}}\right)(\delta_{in} + \delta_{jn})^2\left[1 + \frac{t}{2\sqrt{n}}(\delta_{in} + \delta_{jn})\right]^{-1}.$$

It follows at once that

$$\mu_{ij} = \tfrac{1}{4}t(\delta_i - \delta_j), \tag{47}$$

and

$$\lambda_{ij} = \tfrac{1}{8}t^2(\delta_i - \delta_j)^2. \tag{48}$$

If an overall test of H_0, the equality of all treatment parameters, here called the multi-binomial test, were made on the basis of the $\frac{1}{2}t(t-1)$ independent sets of binomial trials, the appropriate statistic would be

$$S = \sum_{i<j} S_{ij}. \tag{49}$$

From the foregoing argument and the additive property of the χ^2-distribution, under H_0, S has the central χ^2-distribution with $\frac{1}{2}t(t-1)$ degrees of freedom and, under H_a, S has the non-central χ^2-distribution with the same degrees of freedom and parameter of non-centrality,

$$\lambda' = \tfrac{1}{8}t^2\sum_{i<j}(\delta_i - \delta_j)^2 = \tfrac{1}{4}t^2\sum_i \delta_i^2 = \lambda, \tag{50}$$

in the limit with n. The simplification in (50) is possible since $\sum_i \delta_i = 0$.

The paired comparisons test and the multi-binomial test under the conditions appropriate to the former have asymptotic power functions that differ only in the numbers of degrees of freedom that are available. The asymptotic powers of the two tests are clearly different. In Figs. 1 and 2 we have plotted the power β against λ for $t = 3$ and $t = 4$ for both tests based on both 0·05 and 0·01 levels of significance. The β-scale is logarithmic and the power curves were obtained from similar charts given by Pearson & Hartley (1951). The horizontal scales on our charts are for λ, the element common to the parameters of the test considered. To use the Pearson & Hartley charts we computed $\phi = \sqrt{(\lambda/t)}$ for the paired comparisons test as given in (44) and we required

$$\phi' = \sqrt{\frac{\lambda}{\frac{1}{2}t(t-1)+1}} \tag{51}$$

for the multi-binomial test. It is to be noted that, when $\lambda = 0$, $\beta = 0.05$ or 0.01 for both tests. Further, as one would expect, $\beta \to 1$ as $\lambda \to \infty$ for both tests. In the intervening λ-region, the T-test is more powerful than the multi-binomial test, and there is an indication that the advantage of the paired comparisons test increases sharply with t, the number of treatments.

Power curves for an analysis-of-variance test are also given in Figs. 1 and 2. The comparison of paired comparisons with analysis of variance is discussed in the next section.

5·2. *Comparison of the test for paired comparisons with the analysis of variance*

A correspondence between the parameters of the method of paired comparisons and those of analysis of variance may be established by referring to the discussion of Thurstone's model for paired comparisons given by Mosteller (1951). Mosteller summarizes the Thurstone model by writing*

$$P(X_i > X_j) = \frac{1}{\sqrt{(2\pi)}} \int_{-(S_i-S_j)/\sigma_d}^{\infty} e^{-\frac{1}{2}y^2} dy, \tag{52}$$

where S_i is the location point for the variate X_i on a subjective continuum and σ_d^2 is the variance of the difference $(X_i - X_j)$. It was assumed that the variance of X_i is σ^2 and the covariance of X_i and X_j is $\rho\sigma^2$. Then $\sigma_d^2 = 2\sigma^2(1-\rho)$. If X_i and X_j can be observed in paired comparisons, the usual additive model of the analysis of variance would require that $\rho = 0$, S_i be the effect of treatment i, and σ^2 be the common variance of the random elements of that additive model. For our purposes, the Thurstone model may more generally be associated with any analysis of variance with the same variance σ^2 for the random element of its additive model and with S_i as the effect of treatment i.

The present author (1953) has shown that, for the method of paired comparisons here discussed, one can write

$$P(X_i > X_j) = \frac{\pi_i}{\pi_i + \pi_j} = \frac{1}{4} \int_{-(\ln \pi_i - \ln \pi_j)}^{\infty} \operatorname{sech}^2 \tfrac{1}{2}y \, dy = \frac{1}{4a} \int_{-a(\ln \pi_i - \ln \pi_j)}^{\infty} \operatorname{sech}^2 \frac{y}{2a} \, dy. \tag{53}$$

A correspondence between S_i and π_i is required to compare the method of paired comparisons with the analysis of variance. That the forms of the integrand in (52) and (53) do not matter for large-sample sizes can be demonstrated.

---

* Mosteller indicated in his discussion that no generality was lost by taking $\sigma_d = 1$ and he wrote (52) with that restriction.

Consider

$$P(X_i > X_j) = \int_c^\infty f(y)\, dy.$$

The variance of an estimate c_e of c, obtained from an estimate p of $P(X_i > X_j)$ can for large samples be written[*]

$$V(c_e) = \frac{1}{n} P(X_i > X_j)\{1 - P(X_i > X_j)\}\{f(-c)\}^{-2}.$$
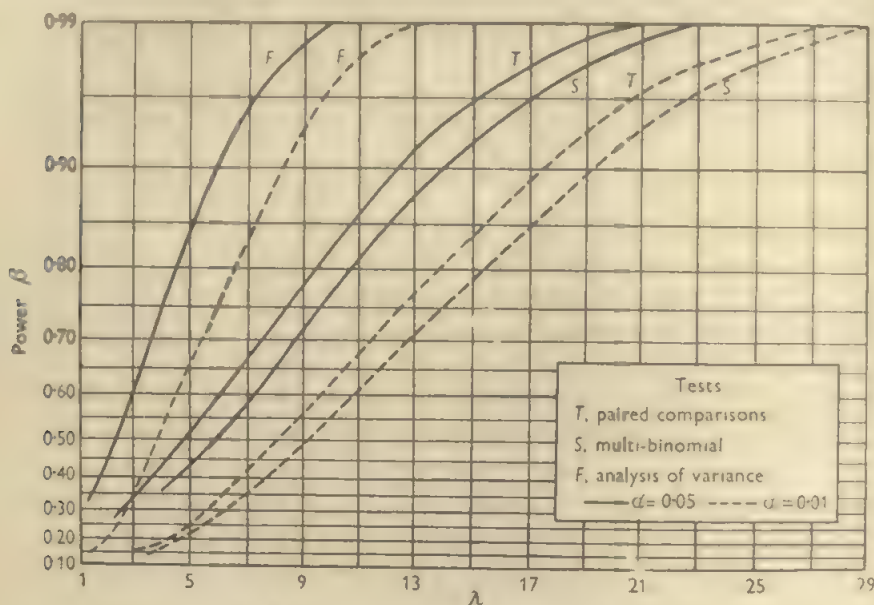
Fig. 1. Asymptotic power functions: $t = 3$.
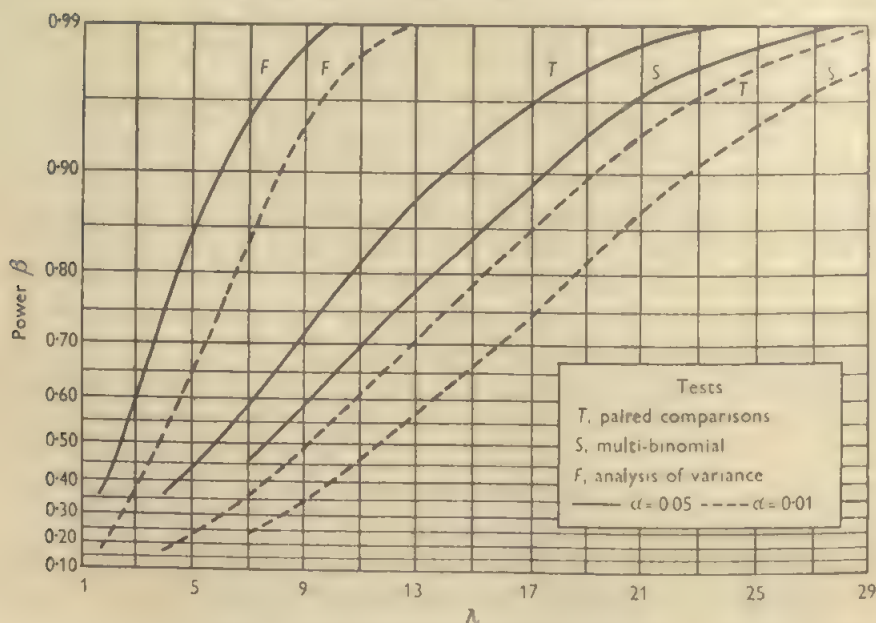
Fig. 2. Asymptotic power functions: $t = 4$.

[*] For a derivation when $f(y)$ is normal, see Cochran (1937). The result is used by Dixon & Mood (1946) in discussing the statistical sign test.

$[\ldots]$ does not depend on the form of $f(s)$ but only on the estimate $[\ldots]$ Accordingly we establish a correspondence between S and g when we adjust the scale factor a in (54) so that the ordinate is equal to $1/\sqrt{(2\pi)}$. The ordinate is $\ldots$ and is evaluated at $g = 0$ since $\ln \pi_i - \ln \pi_j = (\Lambda n^{-1})$ because of the definition $\pi_1 + \ldots = 1$. It follows that

$$a = \sqrt{(\tfrac{1}{2}\pi)} \tag{54}$$

and $\sqrt{(\tfrac{1}{2}\pi)}\ln \pi_i$ corresponds to S_i/σ_d.

If in the Thurstone model we assume that $\mu = 0$ as required for the association with analysis of variance, $\sqrt{2}\,S_i\,\sigma_d$ represents the location point per treatment effect for the ith treatment on a subjective continuum (or additive scale) whereon the variance of an observation is unity. In terms of the parameters of the method of paired comparisons, the location point is $\sqrt{(\tfrac{1}{2}\pi)}\ln \pi_i$. If $n(t-1)$ observations are made on each variate X_i in the analysis of variance (this corresponds to the number of times a treatment appears in the method of paired comparisons), the parameter of non-centrality for the F-test would be approximately

$$\lambda_n'' = \frac{\pi n(t-1)}{4}\sum_i\left(\ln \pi_i - \frac{1}{t}\sum_j \ln \pi_j\right)^2. \tag{55}$$

For large samples the non-central F-distribution approaches the distribution of the non-central χ^2 with $(t-1)$ degrees of freedom and $\lambda_n'' \to \lambda''$ where

$$\lambda'' = \tfrac{1}{2}\pi t^2(t-1)\sum_i \delta_i^2 = \pi(t-1)\lambda/t, \tag{56}$$

in view of (34) and (50).

The power functions for analysis of variance are plotted in Figs 1 and 2. To plot these curves, we computed $\phi'' = \sqrt{[\pi(t-1)\lambda/t^2]}$ in order to use the Pearson & Hartley charts. It is clear that analysis of variance is superior to paired comparisons and the multi-binomial procedure for $t = 3$ and 4 and the advantage increases with t. On the other hand, for comparative purpose, we have assumed that conditions for the valid application of analysis of variance could be attained while the method of paired comparisons was devised for situations where the analysis of variance may not be used.

5·3. Relative efficiencies

The relative efficiency of one test to another may be obtained from the limiting ratio of sample sizes required for equal powers. The concept depends on local properties of the power functions of the two tests being compared. Relative efficiency is the limit of the inverse ratio of sample sizes required in two tests in order that the tests have equal powers for alternatives approaching, with increasing sample sizes, parameter values specified by the null hypothesis. The methods employed in this section parallel those developed by Pitman (1948). We shall first obtain the asymptotic relative efficiency of paired comparisons to analysis of variance by elementary means. Then, to complete the inter-comparisons, we use a more general theorem of Pitman, as presented and extended by Noether (1955), to obtain the relative efficiencies of the multi-binomial procedure to paired comparisons and to the analysis of variance.

Let n'' be the same size for the analysis of variance and n the sample size for paired comparisons. For comparable treatment differences with the ratio n''/n fixed, we need δ_{in}'' corresponding to δ_{in} of (34) such that

$$\lim_{n,\,n''\to\infty}\left\{\sum_i \frac{\delta_{in}''^2}{n''}\Big/\sum_i \frac{\delta_{in}^2}{n}\right\} = 1. \tag{57}$$

Now we have already seen (§ 4·3) that for the method of paired comparisons,

$$\lambda = \tfrac{1}{4} t^3 \sum_i \delta_i^2 = \lim_{n\to\infty} \tfrac{1}{4} n t^3 \sum_i \delta_{in}^2/n,$$

and we can write, from (56),

$$\lambda'' = \lim_{n''\to\infty} \tfrac{1}{4} \pi n'' t^2 (t-1) \sum_i \delta_{in''}''^2/n''.$$

The test situations are comparable when (57) holds and the asymptotic powers are identical when $\lambda = \lambda''$. Consequently the ratio of sample sizes for equal powers is asymptotically

$$\lim_{n,\,n''\to\infty} \frac{n''}{n} = \frac{t}{\pi(t-1)}, \tag{58}$$

and this is the asymptotic relative efficiency of the method of paired comparisons to the analysis of variance.

The theorems of Pitman and Noether yield relative efficiencies in terms of the efficacies of the tests being compared. For a statistic T_N based on a sample of size N, define

$$E(T_N) = \psi_N(\theta; T_N) \quad\text{and}\quad \mathrm{var}\,(T_N) = \sigma_N^2(\theta; T_N).$$

The efficacy of the test of $H_0: \theta = \theta_0$ against the alternative $H: \theta > \theta_0$ is

$$1/R_N^{m\delta}(\theta_0; T_N) = \{\psi^{(m)}(\theta_0; T_N)/\sigma_N(\theta_0; T_N)\}^{1/m\delta} \tag{59}$$

where $\theta_N = \theta_0 + k/N^\delta$. It is assumed that a particular alternative $\theta = \theta_N$ changes with the sample size N in such a way that $\lim_{N\to\infty} \theta_N = \theta_0$.

We have three tests of the hypothesis $H_0: \pi_i = 1/t$ with the alternative,

$$H_a: \quad \pi_i = t^{-1} + n^{-\frac{1}{2}}\delta_{in}, \quad \delta_{in} \to \delta_i \quad\text{as}\quad n\to\infty \quad (i = 1,\ldots,t).$$

The statistics, T_n for the method of paired comparisons, S_n for the multi-binomial procedure, and F_n for the analysis of variance, all have asymptotically non-central χ^2-distributions with parameters,

$$\lambda = \lim_{n\to\infty} \tfrac{1}{4} n t^3 \theta_n, \quad \lambda' = \lim_{n\to\infty} \tfrac{1}{4} n t^3 \theta_n, \quad \lambda'' = \lim_{n\to\infty} \tfrac{1}{4} n \pi t^2 (t-1) \theta_n. \tag{60}$$

and degrees of freedom, $(t-1)$, $\tfrac{1}{2} t(t-1)$, $(t-1)$ respectively. In (60)

$$\theta_n = \sum_i \delta_{in}^2/n. \tag{61}$$

The subscripts were added to the statistics T, S and F in order to provide a notation consistent with the general definition (59) and to emphasize the dependence of the statistics on n.

The hypothesis H_0 is true when $\theta_n = \theta_0 = 0$. We may take $\psi_n(\theta; T_n)$, $\psi_n(\theta; S_n)$ and $\psi_n(\theta; F_n)$, corresponding to $\psi_N(\theta; T_N)$ in (59), respectively to be[*]

$$(t-1) + \tfrac{1}{4} n t^3 \theta, \quad \tfrac{1}{2} t(t-1) + \tfrac{1}{4} n t^3 \theta \quad\text{and}\quad (t-1) + \tfrac{1}{4} n \pi t^2(t-1)\,\theta.$$

Differentiation with respect to θ yields (for all values of θ including $\theta = 0$)

$$\psi_n'(0; T_n) = \tfrac{1}{4} n t^3, \quad \psi_n'(0; S_n) = \tfrac{1}{4} n t^3 \quad\text{and}\quad \psi_n'(0; F_n) = \tfrac{1}{4} n \pi t^2(t-1). \tag{62}$$

---

[*] Our values of ψ_n and σ_n^2, the latter given in (63), are not the means and variances of the statistics for finite n but rather were calculated from the limiting distributions. That this is permissible under conditions which may be verified for our tests was noted by Noether.

From the non-central χ^2-distributions,

$$\sigma_n^2(0; T_n) = 2(t-1), \quad \sigma_n^2(0; S_n) = t(t-1) \quad \text{and} \quad \sigma_n^2(0; F_n) = 2(t-1) \tag{63}$$

The efficacies of the three tests are, with $\delta = 1$, $m = 1$ in (59) in each case,

$$R_n(0; T_n) = \frac{nt^3}{\sqrt{[32(t-1)]}}, \quad R_n(0; S_n) = \frac{nt^2}{4}\sqrt{\frac{t}{t-1}} \quad \text{and} \quad R_n(0; F_n) = \pi n t^2 \sqrt{\frac{t-1}{32}} \tag{64}$$

The relative efficiencies of the three tests taken in pairs are given by the limits of the ratios of corresponding values of R_n subject to conditions set forth in the reference and which may be verified here. The relative efficiencies of S to T and to F are

$$\text{R.E. } (S \text{ to } T) = \lim_{n \to \infty} \frac{R_n(0; S_n)}{R_n(0; T_n)} = \left(\frac{2}{t}\right)^{\frac{1}{2}} \tag{65}$$

and

$$\text{R.E. } (S \text{ to } F) = \lim_{n \to \infty} \frac{R_n(0; S_n)}{R_n(0; F_n)} = \frac{(2t)^{\frac{1}{2}}}{\pi(t-1)}. \tag{66}$$

Table 1. *Asymptotic relative efficiencies*

(*T*, paired comparisons; *S*, multi-binomial; *F*, analysis of variance)

t	2	3	4	5	6	7	8	9	10	∞
T to F	63·7	47·7	42·4	39·8	38·2	37·1	36·4	35·8	35·4	31·8
S to T	100·0	81·7	70·7	63·2	57·7	53·5	50·0	47·1	44·7	0
S to F	63·7	39·0	29·9	25·1	22·0	19·7	18·1	16·9	15·9	0

The result obtained more directly for the relative efficiency of T to F may also be obtained using values given in (64). Asymptotic relative efficiencies expressed as percentages for values of t from 2 to 10 are given in Table 1.

When $t = 2$, the method of paired comparisons reduced to the statistical sign test and the relative efficiency (58) is $2/\pi$ as obtained by Dixon & Mood. The method of paired comparisons becomes more and more inefficient relative to the analysis of variance as t increases. This was to be expected in view of the work of Dixon & Mood for the sign test, and since the comparison was made for a situation in which the analysis of variance could appropriately be used. The assumptions of the analysis of variance are invalidated in much of the work on sensory difference and subjective testing, and it was to avoid those restrictive assumptions that the method of paired comparisons was developed. Further, although for $t = 2$, the tabled relative efficiency of T to F is only 63·7 %, Walsh (1946) has shown that, for samples of size $n = 4$, 5 and 6, the sign test is approximately 95 % as efficient as the 'Student' t-test, and, although the relative efficiency decreases as n increases, it is approximately 75 % when $n = 13$. Dixon & Mood (1946) and Dixon (1953) also indicate that the efficiency of the sign test is better for smaller sample sizes. It seems safe to assume that this situation also holds for $t > 2$ in paired comparisons.

The multi-binomial procedure is seen to have rapidly decreasing relative efficiencies as t increases in comparison with both the method of paired comparisons and the analysis of variance.

VI. ILLUSTRATIVE EXAMPLES

6·1. *Estimated variances and covariances in a numerical example*

We shall use the data from the taste-testing experiment on the flavour characteristics of pork roasts as given as an example by Bradley & Terry (1952). We now assume that the data for the two judges may be pooled. The pork roasts were obtained from hogs fed corn rations and peanut supplements C, Cp and CP. The experiment, with the results of the two judges pooled, yielded the set of sums of ranks, 32, 28 and 30, respectively, for the rations while $n = 10$. The use of the tables in that paper showed the following values:

Σr_1	Σr_2	Σr_3	p_1	p_2	p_3	Prob.
32	28	30	0·24	0·43	0·32	0·63

This result is clearly non-significant at any realistic level of significance and is not indicative of any real ration effect on the flavours of the roasts.

We now obtain estimates of the variances and covariances of $\sqrt{n}\,p_1$, $\sqrt{n}\,p_2$ and $\sqrt{n}\,p_3$ or $\sqrt{n}\,(p_1-\pi_1)$, $\sqrt{n}\,(p_2-\pi_2)$ and $\sqrt{n}\,(p_3-\pi_3)$. The first computing step is to obtain values of $\hat{\lambda}_{ij}$ from substitution of values of p_i for π_i in (14). Then,

$$\hat{\lambda}_{11} = 8\cdot243, \quad \hat{\lambda}_{22} = 2\cdot566,$$
$$\hat{\lambda}_{12} = -2\cdot228, \quad \hat{\lambda}_{23} = -1\cdot778,$$
$$\hat{\lambda}_{13} = -3\cdot189, \quad \hat{\lambda}_{33} = 4\cdot780.$$

The estimate of the determinant in the denominator of (23) is

$$\begin{vmatrix} 8\cdot243 & -2\cdot228 & -3\cdot189 & 1 \\ -2\cdot228 & 2\cdot566 & -1\cdot778 & 1 \\ -3\cdot189 & -1\cdot778 & 4\cdot780 & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix} = -154\cdot976.$$

Now, for example, from (23),

$$\hat{\sigma}_{12} = \frac{1}{154\cdot976} \begin{vmatrix} -2\cdot228 & -1\cdot778 & 1 \\ -3\cdot189 & 4\cdot780 & 1 \\ 1 & 1 & 0 \end{vmatrix} = -0\cdot0485,$$

and similarly, the complete set of estimated variances and covariances is

$$\hat{\sigma}_{11} = 0\cdot0703, \quad \hat{\sigma}_{22} = 0\cdot1252,$$
$$\hat{\sigma}_{12} = -0\cdot0485, \quad \hat{\sigma}_{23} = -0\cdot0767,$$
$$\hat{\sigma}_{13} = -0\cdot0218, \quad \hat{\sigma}_{33} = 0\cdot0985.$$

The estimated variances and covariances of p_1, p_2 and p_3 may be obtained by dividing those immediately above by $n = 10$. Consequently, the estimated standard errors of p_1, p_2 and p_3 are respectively given approximately by 0·084, 0·112 and 0·099.

A check on the computing may be made by computing the variance of $\sqrt{n}\,\sum_i p_i$ in terms of the variances and covariances of the elements of this sum. The result of course should be zero.

6·2. *Use of the power function*

(i) Suppose that the true values of the parameters π_1, π_2 and π_3 in the example of the preceding section are respectively 0·28, 0·59 and 0·13. (If these were estimates with $n = 10$, the significance level of the test would be 0·04.) We use δ_{iu} defined in (34) as an approximation to δ_i and from (45) and (44) we have the approximate values

$$\lambda = 7\cdot43 \quad \text{and} \quad \phi = 1\cdot57.$$

The approximate power of a 0·05-level test, obtained from Tang's tables* with $f_1 = 2$, $f_2 = \infty$, $\alpha = 0\cdot05$, is $\beta(7\cdot43 \mid 0\cdot05, 2, \infty) = 0\cdot67$ and with $\alpha = 0\cdot01$, $\beta(7\cdot43 \mid 0\cdot01, 2, \infty) = 0\cdot56$ in the notation of (43). If we are interested in the probability of failing to reject H_0 when H_1 is true, we require $1 - \beta$. For $\alpha = 0\cdot05$, the probability of a type II error is 0·33 and, for $\alpha = 0\cdot01$, the probability of a type II error is 0·44.

(ii) Suppose that the differences observed in the estimates for the pork roast experiment are of practical importance (as distinct from statistical significance) and that for a 0·05-level test we desire the sample size for a second follow-up experiment such that $\beta = 0\cdot95$ or such that the probability of a type II error will be about 0·05. Without specifying n, we take the values of π_1, π_2 and π_3 to be 0·24, 0·43 and 0·32 as estimated in the experiment. Then from (45), $\lambda = 0\cdot123n$ and $\phi = \sqrt{(0\cdot123n/3)}$.

To determine n, we enter Tang's tables with $f_1 = 2$, $f_2 = \infty$, $\alpha = 0\cdot05$ and find that ϕ must have the value 2·35 for $\beta(0\cdot123n \mid 0\cdot05, 2, \infty) = 0\cdot95$. The number of repetitions required is obtained by setting

$$0\cdot123n/3 = (2\cdot35)^2,$$

from whence

$$n = 135.$$

This is of course a very large number of repetitions and is required owing to the small differences among the estimates of the experiment.

If we take the hypothetical case considered in (i) above and the same test requirements,

$$0\cdot743n/3 = (2\cdot35)^2 \quad \text{and} \quad n = 22.$$

This value may again seem high but it must be emphasized that most experiments are conducted without regard to the powers of the tests employed and this is true even when the analysis of variance is used.

VII. DISCUSSION AND SUMMARY

7·1. *Additional test procedures*

Two test procedures relating to paired comparisons were discussed by Bradley & Terry (1952) and have not yet been considered in this paper.

The first test is a combined test on the equality of treatment effects with the specification of the alternative hypothesis such that the parameters may differ from one group of repetitions to another. Suppose the paired comparisons experiment is performed in g groups of repetitions with n_u repetitions in the uth group, $\sum_{u=1}^{g} n_u = n$, and suppose that the treatments have parameters, $\pi_{1u}, \ldots, \pi_{tu}$, $\sum_i \pi_{iu} = 1$, in the uth group. The hypotheses are

$$H_0: \quad \pi_{iu} = 1/t \quad (i = 1, \ldots, t; \; u = 1, \ldots, g)$$

* Tang (1938) tabulated P_{II}, the probability of a type II error. In using the tables, linear interpolation was judged to be satisfactory for our examples.

and H_a: No π_{iu} assumed equal to any other π_{jv} $(i \neq j$ or $u \neq v;\ i,j = 1, \ldots, t;\ u, v = 1, \ldots, g)$ Let T_u be the statistic corresponding to T for the uth group. Then

$$T_c = \sum_{u=1}^{g} T_u$$

is the appropriate statistic for this test. From the additive property of χ^2, it follows that under H_0, T_c has, for large n and fixed ratios n_u/n, the central χ^2-distribution with $g(t-1)$ degrees of freedom and, under H_a, T_c has the non-central χ^2-distribution with $g(t-1)$ degrees of freedom and parameter of non-centrality approximated by

$$\lambda_c = \tfrac{1}{4} t^3 \sum_i \sum_u^g n_u \left(\pi_{iu} - \frac{1}{t} \right)^2.$$

The second test is a test of agreement of ranking from group to group. The hypotheses are H_0: $\pi_{iu} = \pi_i$ $(i = 1, \ldots, t;\ u = 1, \ldots, g)$ and H_a as specified in the preceding paragraph. Let T_p be the T-statistic obtained by pooling all groups of repetitions as though we have a homogeneous set of n repetitions. The appropriate statistic for this second test is $(T_c - T_p)$. This statistic has the central χ^2-distribution with $(g-1)(t-1)$ degrees of freedom under H_0. Under H_a, we state without proof that the distribution of $(T_c - T_p)$ is the non-central χ^2-distribution with $(g-1)(t-1)$ degrees of freedom and parameter of non-centrality approximated by

$$\lambda_a = \tfrac{1}{4} t^3 \sum_i \sum_{u=1}^{g} n_u (\pi_{iu}^2 - \pi_i^2).$$

7·2. *Discussion*

The results obtained in this paper are valid only for large samples. For small samples they may be approximately correct, but the difficulty of obtaining exact powers makes a check very nearly impossible. Under H_0 some information is available on the approach of the distribution of T to the central χ^2-distribution. The means and variances of T were computed from the exact small-sample distribution of T (Bradley, 1954b) and their approach to the corresponding moments of χ^2 noted. It appeared that the approximate tests would be adequate at least for $n \geqslant 10$.

In extreme cases, the estimators $p_1, \ldots, p_t$ sometimes have the set of values $1, 0, \ldots, 0$.

This prohibits the computation of $\hat{\lambda}_{ij}$ and the estimation of variances and covariances of the estimators. This situation was previously discussed (Bradley, 1954b), and it was noted that the first treatment could be eliminated from the analysis and that secondary parameter estimates could be obtained for $\pi_2, \ldots, \pi_t$. This procedure again seems suitable when the specified set of values occurs. Then it will be possible to proceed with the analysis and the estimation of variances and covariances for the remaining estimates. Other extreme sets of sums of ranks yield extreme sets of estimators and methods of dealing with them are discussed in the cited reference.

The comparisons of the properties of the method of paired comparisons and of other test methods produce results that seem to be about what one would expect. The method of paired comparisons is clearly superior to the multi-binomial procedure both on the basis of the asymptotic powers of the tests for equal sample sizes and the study of their relative efficiency. It appears that the superiority of paired comparisons will become more marked as the number of treatments is increased. The comparison of the method of paired comparisons with the analysis of variance is less favourable to the method of paired comparisons.

Here the comparison was made when the analysis of variance would be the appropriate method. It should be remembered that the method of paired comparisons was formulated largely for subjective tests and it is there that the assumptions of the analysis of variance are seriously suspect. Further, even if a scoring technique could be devised for which the analysis of variance could be used, it is quite possible that an over-all consideration of time of judging and analysis would dictate the use of more samples and the non-parametric method. The large-sample results indicate that for the usual numbers of treatments $2\frac{1}{2}$ times as many samples are required for the method of paired comparisons as for the analysis of variance. It is conceivable in a tasting experiment, for example, that an individual can indicate rankings in a paired-treatment experiment as contrasted to scoring individual samples on an 11-point scale, say, with much less taste fatigue and with a speed ratio in favour of ranking in excess of the ratio of sample numbers.

7·3. *Summary*

In this paper we have examined some of the properties of the method of paired comparisons. The results obtained are asymptotically correct for large-sample sizes or for large numbers of repetitions of the paired-comparisons design. Formulae for the variances and covariances of estimates of treatment ratings, $\pi_1, \ldots, \pi_t$, have been obtained, and these were not heretofore available. It has been shown that statistics previously used for tests of significance have limiting non-central χ^2-distributions, and the appropriate parameters of non-centrality are given.

It was found that in comparison with the analysis of variance the relative efficiency of the method of paired comparisons is $t/\{\pi(t-1)\}$, and, when $t = 2$, this has the value $2/\pi$ previously obtained in comparing the sign test with the 'Student' t-test. The method of paired comparisons was seen to be considerably better than a multi-binomial procedure postulated and the asymptotic powers of these two tests are plotted for a number of examples along with similar values for the analysis of variance.

Examples of the use of the power function developed were given in application to an experiment in taste testing. Estimated variances and covariances of the estimators of the example were computed.

REFERENCES

ABELSON, R. M. & BRADLEY, R. A. (1954). A 2×2 factorial with paired comparisons. *Biometrics*, 10, 487.

BRADLEY, R. A. (1953). Some statistical methods in taste testing and quality evaluation. *Biometrics*, 9, 22.

BRADLEY, R. A. (1954a). Incomplete block rank analysis: on the appropriateness of the model for a method of paired comparisons. *Biometrics*, 10, 375.

BRADLEY, R. A. (1954b). Rank analysis of incomplete block designs. II. Additional tables for the method of paired comparisons. *Biometrika*, 41, 502.

BRADLEY, R. A. & TERRY, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, 39, 324.

CHANDA, K. C. (1954). A note on the consistency and maxima of the roots of likelihood equations. *Biometrika*, 41, 56.

COCHRAN, W. G. (1937). The efficiencies of the binomial series test of significance of a mean and of a correlation coefficient. *J. R. Statist. Soc.* A, 100, 69.

CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

DIXON, W. J. (1953). Power functions of the sign test and power efficiency for normal alternatives. *Ann. Math. Statist.* 24, 467.

DIXON, W. J. & MOOD, A. M. (1946). The statistical sign test. *J. Amer. Statist. Ass.* 41, 557.

FIX, EVELYN (1949). Tables of noncentral χ^2. *Univ. Calif. Publ. Statist.* 1 (2), 15.

HALD, A. (1952). *Statistical Theory with Engineering Applications*, §9·9. New York: John Wiley and Sons, Inc.

HEMELRIJK, J. (1952). A theorem on the sign test when ties are present. *Indag. Math.* 14, 322.

HOPKINS, J. W. (1954). Incomplete block rank analysis: some taste test results. *Biometrics*, 10, 391.

MOSTELLER, F. (1951). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. *Psychometrika*, 16, 3.

NOETHER, G. (1955). On a theorem of Pitman. *Ann. Math. Statist.* 26, 64.

PEARSON, E. S. & HARTLEY, H. O. (1951). Charts for the power function for analysis of variance tests, revised from the non-central *F*-distribution. *Biometrika*, 38, 112.

PITMAN, E. J. G. (1948). *Lecture Notes on Non-Parametric Statistics*. Columbia University, New York.

TANG, P. C. (1938). The power function of analysis of variance tests with tables and illustrations of their use. *Statist. Res. Mem.* 2, 126.

TERRY, M. E., BRADLEY, R. A. & DAVIS, L. L. (1952). New designs and techniques for organoleptic testing. *Food Tech., Lond.*, 6, 250.

WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* 54, 426.

WALSH, J. (1946). On the power function of the sign test for slippage of means. *Ann. Math. Statist.* 17, 358.

A METHOD OF ASSIGNING CONFIDENCE LIMITS TO LINEAR COMBINATIONS OF VARIANCES

By A. HUITSON, Ph.D.

Royal Naval Scientific Service

1. Introduction and summary

Linear combinations of variances occur frequently in analysis of variance work, when we are considering the estimation either of individual components of variance or of gross variability. Percentage points for such functions cannot be computed directly in the same manner as that adopted when the χ^2 distribution is used to give percentage points in problems where only a single unknown variance is involved.

It is the purpose of the present paper (a) to derive a series approximation to the percentage points of functions of this type, (b) to investigate the numerical behaviour of the expansion in some particular cases and (c) to provide tables of the series expansion, suitable for use in certain practical problems. An example is given in which use is made of the tables, together with some general discussion of the single-factor type of experiment.

2. Development of a series approximation

Let us consider any number of unknown population variances which we shall denote by $\sigma_1^2, \sigma_2^2, \ldots, \sigma_r^2$.*

Suppose that the observed data provide estimates s_i^2 of these variances based on f_i degrees of freedom respectively, and that these estimates are distributed independently of each other in the form

$$p(s_i^2)\,ds_i^2 = \frac{1}{\Gamma(\tfrac{1}{2}f_i)} \left(\frac{f_i s_i^2}{2\sigma_i^2}\right)^{\frac{1}{2}f_i-1} \exp\left[-\frac{1}{2}\frac{f_i s_i^2}{\sigma_i^2}\right] d\left(\frac{f_i s_i^2}{2\sigma_i^2}\right). \tag{1}$$

We are seeking to assign confidence limits, calculable from the observations, to the expression $\sum_{i=1}^{r} \lambda_i \sigma_i^2$, where $\lambda_1 \ldots \lambda_r$ are known arbitrary constants. In order to accomplish this, we shall seek a function y, which will satisfy the equation

$$P_r\left[\frac{\sum_{i=1}^{r}\lambda_i s_i^2}{\sum_{i=1}^{r}\lambda_i \sigma_i^2} < y(s_1^2, s_2^2, \ldots, s_r^2, P)\right] = P, \tag{2}$$

where P_r is used to denote 'the probability of the relation in the bracket following'. It is by no means obvious that there is a function y having the required property, but as we shall see later it is certainly possible to find various functions which have *approximately* the required properties. We shall be content to leave open the question as to whether our problem may be solved exactly.

---

* More strictly, in most practical applications, the $\sigma_1^2, \sigma_2^2, \ldots, \sigma_r^2$ are parameters which are already linear functions of more basic variances. It will be convenient, however, to refer to the $\sigma_1^2, \sigma_2^2, \ldots, \sigma_r^2$ themselves as variances since they have estimates $s_1^2, s_2^2, \ldots, s_r^2$ which are distributed in the form (1).

The large-sample normal approximation to y is

$$1 + \xi \frac{\sqrt{2}}{(\Sigma \lambda_i \sigma_i^2)} \sqrt{\left(\Sigma \lambda_i^{\frac{1}{2}} \frac{\sigma_i^4}{f_i} \right)},$$

where ξ represents the Pth percentage point of the unit normal deviate. However, as the variances are unknown, it will be convenient to take as our initial approximation to y

$$1 + \xi \frac{\sqrt{2}}{(\Sigma \lambda_i s_i^2)} \sqrt{\left(\Sigma \lambda_i^{\frac{1}{2}} \frac{s_i^4}{f_i} \right)}. \qquad (3)$$

All summations are to be taken over the range 1 to r.

We shall now introduce the notation

$$V_{mn} = \frac{\Sigma \dfrac{\lambda_i^m s_i^{2m}}{f_i^n}}{(\Sigma \lambda_i s_i^2)^m} = \Sigma \frac{c_i^m}{f_i^n}, \quad \text{where} \quad c_i = \frac{\lambda_i s_i^2}{(\Sigma \lambda_i s_i^2)},$$

and denote the corresponding functions of the unknown variances by Δ_{mn} and γ_i respectively. Thus our initial approximation to y becomes

$$1 + \xi \sqrt{(2V_{21})}. \qquad (4)$$

If we assume that y can be written in the form

$$y = h_0 + h_1 + h_2 + h_3 + \dots, \qquad (5)$$

where h_i is of order $f^{-\frac{1}{2}i}$, then the large-sample normal approximation implies that

$$\left. \begin{array}{l} h_0 = 1, \\ h_1 = \xi \sqrt{(2V_{21})}. \end{array} \right\} \qquad (6)$$

We shall now proceed to derive the next term, h_2, in the series expansion of y. Let us consider the new variable

$$u = \frac{\Sigma \lambda_i s_i^2}{\Sigma \lambda_i \sigma_i^2} - (h_0 + h_1). \qquad (7)$$

As u is a function of the variance estimates, the moment-generating function of u may be found by averaging over the distributions of the s_i^2. On simplification we find the first three cumulants of u to be

$$K_1 = -\xi \sqrt{(2\Delta_{21})} \quad \text{to order } f^{-1},$$
$$K_2 = 2\Delta_{21} + 4\sqrt{2} \xi \{(\Delta_{21})^{\frac{3}{2}} - \Delta_{32}(\Delta_{21})^{-\frac{1}{2}}\} \quad \text{to order } f^{-\frac{3}{2}},$$

and
$$K_3 = 8\Delta_{32} \quad \text{to order } f^{-2}.$$

Thus we can see that $K_1 = O(f^{-\frac{1}{2}})$, $K_2 = O(f^{-1})$ and $K_3 = O(f^{-2})$. It appears that u has cumulants which decrease in powers of f such that $K_r = O(f^{-r+1})$.

Using the Cornish & Fisher expansion (1937) we then obtain after simplification the Pth percentage point of u, u_P say, to be equal to

$$\xi^2 \left[2\Delta_{21} - \frac{4\Delta_{32}}{3\Delta_{21}} \right] - \frac{2\Delta_{32}}{3\Delta_{21}} \quad \text{to order } \left(\frac{1}{f} \right). \qquad (8)$$

But this is a function of the ratios of the unknown variances and thus is not of immediate use to us. Consider any particular ratio, say $\sigma_m^2 \cdot \sigma_n^2$, where σ_m^2 and σ_n^2 are any two of the $\sigma_1^2 \ldots \sigma_r^2$. The mean of s_m^2 / s_n^2 is equal to

$$\frac{\sigma_m^2}{\sigma_n^2}\left(\frac{f_n}{f_n - 2}\right) = \frac{\sigma_m^2}{\sigma_n^2}\left(1 + \frac{2}{f_n} + \ldots\right).$$

Also the variance of $s_m^2 \cdot s_n^2$ is of the order of f^{-1}. Therefore neglecting terms of order f^{-1} and lower we have in the probability sense $s_m^2 \cdot s_n^2 = \sigma_m^2 \cdot \sigma_n^2$. Thus, without affecting the accuracy of equation (8) we can replace all the ratios of the unknown variances by the ratios of their estimates, since all terms on the right-hand side of (8) are of the same order.

Therefore Δ_{mn} will now become V_{mn} and

$$h_2 = \xi^3\left[2V_{21} - \frac{4}{3}\frac{V_{32}}{V_{21}}\right] - \frac{2}{3}\frac{V_{32}}{V_{21}}. \tag{9}$$

The fact that (9) is of order f^{-1} and not of higher order is a verification that we have taken the correct value for h_1. Thus the method may be regarded as self checking, in that, in each application, it checks the value which has been obtained previously and adds another term to the series. Using the same method the next term of the series expansion for y, h_3, has been found to be

$$h_3 = \sqrt{(2V_{21})}\{\xi[-\tfrac{4}{3}V_{32}(V_{21})^{-1} + \tfrac{9}{2}V_{43}(V_{21})^{-2} - V_{22}(V_{21})^{-1} - \tfrac{23}{9}(V_{32})^2(V_{21})^{-3}]$$
$$+ \xi^3[-\tfrac{8}{3}V_{32}(V_{21})^{-1} + 2V_{21} - \tfrac{16}{9}V_{32}^2(V_{21})^{-3} + \tfrac{5}{2}V_{43}(V_{21})^{-2}]\}. \tag{10}$$

However, the solution was carried no further as the work involved would have been too laborious.

3. CHECKS

The following algebraic checks have been carried out on the series expansion of y derived above.

(a) It has been shown to agree with an independent development of the series solution carried through by Welch (1956) for the same problem, although his treatment and the algebra involved are somewhat different in detail from mine.

(b) It has also been shown to agree with a result published by Bartlett (1953). In this paper Bartlett derives a confidence interval for a function of the form $\sigma_1^2 + \lambda\sigma_2^2$. To the appropriate order of terms, $h_0 + h_1 + h_2$ is the solution of this equation.

(c) A simple check, which has also been carried out, is to consider the special case when all the degrees of freedom except one are infinite. A series expansion for the function $\Sigma\lambda_i\sigma_i^2$ can then be found from the χ^2 distribution and this compared term by term with the series derived above.

4. NUMERICAL TESTING

The series approximation given by $h_0 + h_1 + h_2 + h_3$ is applicable to the general problem involving r variances. In order to carry out a numerical investigation of it, we shall limit ourselves to the particular case $r = 2$ only and we shall assume both λ's have the same sign.

We can now write

$$V_{mn} = \frac{c^m}{f_1^n} + \frac{(1-c)^m}{f_2^n}, \quad \text{where} \quad c = \frac{\lambda_1 s_1^2}{\lambda_1 s_1^2 + \lambda_2 s_2^2} \quad \text{and} \quad \gamma = \frac{\lambda_1\sigma_1^2}{\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2}.$$

For the cases in which both the λ's are of the same sign, the permissible values of c lie between 0 and 1.

Consider an approximation to the Pth percentage point of $(\lambda_1 s_1^2 + \lambda_2 s_2^2)$ $(\lambda_1 \sigma_1^2 - \lambda_2 \sigma_2^2)$ depending only on the sample statistic c (it will also depend on the probability P), $g(c)$ say. To test the accuracy of this approximation, we must evaluate

$$P_r\left[\frac{\lambda_1 s_1^2 + \lambda_2 s_2^2}{\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2} < g(c)\right] = Q \quad \text{say}, \tag{11}$$

in order to see how close it is to P. Strictly, such a function as $g(c)$ is not an approximation to the Pth percentage point of $(\lambda_1 s_1^2 + \lambda_2 s_2^2)$ $(\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2)$, as this is a function of the unknown variances while $g(c)$ is independent of these variances. It can only be regarded as such an approximation in the sense that it makes the value of Q in equation (11) close to P.

It can be shown by straightforward transformations that

$$Q = \int_0^1 \frac{I}{f_1 + f_2}\left[\left(\frac{\gamma b}{f_1} + \frac{(1-\gamma)(1-b)}{f_2}\right)^{-1} g\left\{\frac{f_2 \gamma b}{f_2 \gamma b + f_1(1-\gamma)(1-b)}\right\}\right] p(b)\, db. \tag{12}$$

where

$$I[x] = \int_0^x \frac{1}{\Gamma(\frac{1}{2}F)} (\tfrac{1}{2}y)^{\frac{1}{2}F-1} e^{-\frac{1}{2}y} d(\tfrac{1}{2}y),$$

$$p(b)\, db = \frac{1}{B(\frac{1}{2}f_1, \frac{1}{2}f_2)} b^{\frac{1}{2}f_1-1}(1-b)^{\frac{1}{2}f_2-1} db$$

and

$$b = \frac{f_1 s_1^2/\sigma_1^2}{f_1 s_1^2/\sigma_1^2 + f_2 s_2^2/\sigma_2^2}.$$

From equation (12) the following values of Q were obtained by quadrature when $f_1 = f_2 = 10$ with $g(c)$ replaced by the series approximation, cut off at various points, in order to see what effect each successive term of the series had on the result:

Values of Q when $f_1 = f_2 = 10$

	$P = 0.05$		$P = 0.95$	
	$\gamma = 0.5$	$\gamma = 0.2$	$\gamma = 0.5$	$\gamma = 0.2$
$g(c) = h_0 + h_1$	0.0197	0.0269	0.9424	0.9419
$g(c) = h_0 + h_1 + h_2$	0.0358	0.0469	0.9547	0.9533
$g(c) = h_0 + h_1 + h_2 + h_3$	0.0453	0.0544	0.9497	0.9501

The values of Q for $\gamma = \alpha$, say, and $\gamma = 1-\alpha$ will be identical when $f_1 = f_2$. Therefore the values of Q when $\gamma = 0.8$ will correspond to those given in the third and last columns of the above table. In all except the third column it is seen that the addition of each successive term to the series approximation brings the value of Q closer to the chosen value of P. In the third column the addition of h_3 has moved Q in the right direction, but too far.

It may also be noted that the values of Q obtained when $P = 0.05$ are much more in error than the corresponding ones for $P = 0.95$. This suggests that the series approximation may not be so good at the lower end of the distribution as it is at the upper end.

This comparison relates only to the case $r = 2$ and λ_1 and λ_2 positive. More work would be necessary to assess the merits of the series for the more general situation. The results for the case considered are, however, sufficiently promising to make it appear worth while to table some results for the larger numbers of degrees of freedom at least. Such tables are presented in the next section.

5. Tables

A two decimal table will usually be quite adequate for a working statistician. Four probability levels are tabled in this paper (Tables 1-4).

We shall regard the series solution as giving us two decimal accuracy when $h_3 < 0.005$. The argument here is that if h_3 is negligible, then so, almost certainly, are higher terms of the series. If h_3 is not negligible it does not follow that the higher terms are not negligible, but we have no absolute justification for assuming that they are. Using this criterion, the tabled values will be correct, except possibly those for which one of the degrees of freedom is 16, which may be one or two units in error in the second decimal place.

The values of degrees of freedom for which the tables have been constructed are such that their square roots form a harmonic set. The purpose of this is to make interpolation with respect to the degrees of freedom easier. Thus, for intermediate values of f, it is necessary to interpolate with respect to $12/\sqrt{f}$. Tables 1-4 can be used to assign 90 and 98 % confidence intervals to any sum of two variances, provided the degrees of freedom of the variance estimates are large anough.

6. Example

The following example is given by Tippett (1952, p. 111). It is of the single-factor analysis type. Mule cops (bobbins) of cotton yarn were collected in blocks of 20, each block being from a different mule. Two leas (1 lea = 120 yards) were weighed from each cop, giving for each block 39 total degrees of freedom, 19 'between cops' and 20 'within cops', together with corresponding sums of squares. There were six such blocks and the six sets of sums of squares and degrees of freedom were added to give the following table:

Source of variation	Sum of squares	Degrees of freedom	Mean square	Expected value of mean square
Between cops (within blocks)	19,138·85	114	$167.88 = s_1^2$	$\sigma_1^2 = 2\sigma_K^2 + \sigma_I^2$
Within cops	5,681·00	120	$47.34 = s_2^2$	$\sigma_2^2 = \sigma_I^2$
Total	24,819·85	234		

where σ_K^2 is the variance between cops and σ_I^2 is the interaction (or error) variance in the usual manner.

The mean squares give an analysis of the variations within the blocks, and these are interesting because they are due to factors not easily controllable. The block to block variations are not studied because they can be easily eliminated by careful adjustment of the mules.

Table 1. *Lower* 5 % *critical values of* $(\lambda_1 s_1^2 - \lambda_2 s_2^2)(\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2)$

$\dfrac{\lambda_1 s_1^2}{\lambda_1 s_1^2 + \lambda_2 s_2^2}$		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
f_1	f_2											
16	16	0.50	0.52	0.55	0.59	0.62	0.64	0.62	0.59	0.55	0.52	0.50
	36	0.65	0.67	0.69	0.71	0.70	0.66	0.62	0.58	0.55	0.52	0.50
	144	0.81	0.82	0.81	0.76	0.71	0.66	0.62	0.58	0.55	0.52	0.50
	∞	1.00	0.91	0.83	0.77	0.71	0.66	0.62	0.58	0.55	0.52	0.50
36	16	0.50	0.52	0.55	0.58	0.62	0.66	0.70	0.71	0.69	0.67	0.65
	36	0.65	0.67	0.69	0.72	0.74	0.75	0.74	0.72	0.69	0.67	0.65
	144	0.81	0.83	0.83	0.83	0.81	0.78	0.75	0.72	0.69	0.67	0.65
	∞	1.00	0.95	0.90	0.86	0.82	0.78	0.75	0.72	0.70	0.67	0.65
144	16	0.50	0.52	0.55	0.58	0.62	0.66	0.71	0.76	0.81	0.82	0.81
	36	0.65	0.67	0.69	0.72	0.75	0.78	0.81	0.83	0.83	0.83	0.81
	144	0.81	0.83	0.84	0.86	0.86	0.87	0.86	0.86	0.84	0.83	0.81
	∞	1.00	0.98	0.96	0.94	0.92	0.90	0.88	0.86	0.85	0.83	0.81
∞	16	0.50	0.52	0.55	0.58	0.62	0.66	0.71	0.77	0.83	0.91	1.00
	36	0.65	0.67	0.70	0.72	0.75	0.78	0.82	0.86	0.90	0.95	1.00
	144	0.81	0.83	0.85	0.86	0.88	0.90	0.92	0.94	0.96	0.98	1.00
	∞	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 2. *Upper* 5 % *critical values of* $(\lambda_1 s_1^2 + \lambda_2 s_2^2)/(\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2)$

$\dfrac{\lambda_1 s_1^2}{\lambda_1 s_1^2 + \lambda_2 s_2^2}$		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
f_1	f_2											
16	16	1.64	1.56	1.50	1.46	1.44	1.43	1.44	1.46	1.50	1.56	1.64
	36	1.42	1.37	1.35	1.34	1.34	1.36	1.39	1.43	1.48	1.55	1.64
	144	1.20	1.19	1.20	1.22	1.25	1.29	1.33	1.39	1.46	1.55	1.64
	∞	1.00	1.04	1.09	1.13	1.18	1.24	1.30	1.37	1.45	1.54	1.64
36	16	1.64	1.55	1.48	1.43	1.39	1.36	1.34	1.34	1.35	1.37	1.42
	36	1.42	1.37	1.33	1.31	1.29	1.29	1.29	1.31	1.33	1.37	1.42
	144	1.20	1.18	1.18	1.18	1.20	1.21	1.24	1.27	1.31	1.36	1.42
	∞	1.00	1.03	1.06	1.10	1.13	1.17	1.21	1.26	1.31	1.36	1.42
144	16	1.64	1.55	1.46	1.39	1.33	1.29	1.25	1.22	1.20	1.19	1.20
	36	1.42	1.36	1.31	1.27	1.24	1.21	1.20	1.18	1.18	1.18	1.20
	144	1.20	1.18	1.16	1.15	1.14	1.14	1.14	1.15	1.16	1.18	1.20
	∞	1.00	1.02	1.03	1.05	1.07	1.09	1.11	1.13	1.15	1.18	1.20
∞	16	1.64	1.54	1.45	1.37	1.30	1.24	1.18	1.13	1.09	1.04	1.00
	36	1.42	1.36	1.31	1.26	1.21	1.17	1.13	1.10	1.06	1.03	1.00
	144	1.20	1.18	1.15	1.13	1.11	1.09	1.07	1.05	1.03	1.02	1.00
	∞	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

—

Table 3. *Lower 1% critical values of* $(\lambda_1 s_1^2 + \lambda_2 s_2^2)/(\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2)$

$\dfrac{\lambda_1 s_1^2}{\lambda_1 s_1^2 + \lambda_2 s_2^2}$		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
f_1	f_2											
16	16	0.37	0.39	0.41	0.45	0.50	0.53	0.50	0.45	0.41	0.39	0.37
	36	0.53	0.56	0.59	0.61	0.59	0.54	0.48	0.44	0.41	0.39	0.37
	144	0.75	0.76	0.74	0.66	0.58	0.52	0.47	0.44	0.41	0.39	0.37
	∞	1.00	0.86	0.74	0.65	0.58	0.52	0.48	0.44	0.42	0.39	0.37
36	16	0.37	0.39	0.41	0.44	0.48	0.54	0.59	0.61	0.59	0.56	0.53
	36	0.53	0.56	0.59	0.62	0.65	0.66	0.65	0.62	0.59	0.56	0.53
	144	0.75	0.76	0.77	0.76	0.73	0.69	0.65	0.62	0.59	0.56	0.53
	∞	1.00	0.92	0.85	0.79	0.74	0.69	0.65	0.62	0.59	0.56	0.53
144	16	0.37	0.39	0.41	0.44	0.47	0.52	0.58	0.66	0.74	0.76	0.75
	36	0.53	0.56	0.59	0.62	0.65	0.69	0.73	0.76	0.77	0.76	0.75
	144	0.75	0.77	0.78	0.80	0.81	0.82	0.81	0.80	0.78	0.77	0.75
	∞	1.00	0.97	0.94	0.91	0.88	0.85	0.83	0.81	0.79	0.77	0.75
∞	16	0.37	0.39	0.42	0.44	0.48	0.52	0.58	0.65	0.74	0.86	1.00
	36	0.53	0.56	0.59	0.62	0.65	0.69	0.74	0.79	0.85	0.92	1.00
	144	0.75	0.77	0.79	0.81	0.83	0.85	0.88	0.91	0.94	0.97	1.00
	∞	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4. *Upper 1% critical values of* $(\lambda_1 s_1^2 + \lambda_2 s_2^2)/(\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2)$

$\dfrac{\lambda_1 s_1^2}{\lambda_1 s_1^2 + \lambda_2 s_2^2}$		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
f_1	f_2											
16	16	2.00	1.85	1.76	1.70	1.67	1.65	1.67	1.70	1.76	1.85	2.00
	36	1.63	1.55	1.52	1.50	1.51	1.54	1.58	1.63	1.71	1.83	2.00
	144	1.29	1.28	1.29	1.32	1.36	1.41	1.47	1.56	1.68	1.82	2.00
	∞	1.00	1.06	1.11	1.18	1.25	1.33	1.42	1.53	1.66	1.82	2.00
36	16	2.00	1.83	1.71	1.63	1.58	1.54	1.51	1.50	1.52	1.55	1.63
	36	1.63	1.54	1.49	1.45	1.43	1.42	1.43	1.45	1.49	1.54	1.63
	144	1.29	1.27	1.26	1.27	1.28	1.31	1.35	1.40	1.46	1.54	1.63
	∞	1.00	1.04	1.08	1.13	1.18	1.24	1.30	1.37	1.45	1.53	1.63
144	16	2.00	1.82	1.68	1.56	1.47	1.41	1.36	1.32	1.29	1.28	1.29
	36	1.63	1.54	1.46	1.40	1.35	1.31	1.28	1.27	1.26	1.27	1.29
	144	1.29	1.26	1.24	1.22	1.21	1.20	1.21	1.22	1.24	1.26	1.29
	∞	1.00	1.02	1.05	1.07	1.10	1.13	1.16	1.19	1.22	1.26	1.29
∞	16	2.00	1.82	1.66	1.53	1.42	1.33	1.25	1.18	1.11	1.06	1.00
	36	1.63	1.53	1.45	1.37	1.30	1.24	1.18	1.13	1.08	1.04	1.00
	144	1.29	1.26	1.22	1.19	1.16	1.13	1.10	1.07	1.05	1.02	1.00
	∞	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

'Between cops' corresponds to the factor and 'within cops' to the error; but it is inappropriate to speak of error when the investigation is not a controlled, or partially controlled, experiment.

Suppose that we wish to find confidence limits for the gross variability (i.e. the variance of a single observation). The estimate of the gross variability, $\sigma_E^2 + \sigma_I^2 = \frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2$, is

$$\lambda_1 s_1^2 + \lambda_2 s_2^2 = \frac{1}{2}(47\cdot34) + \frac{1}{2}(167\cdot88) = 107\cdot61.$$

Then $c = \lambda_1 s_1^2 (\lambda_1 s_1^2 + \lambda_2 s_2^2) = 0\cdot2200,$ $f_1 = 120,$ $f_2 = 114.$

Interpolating in Table 1 we get a value of $0\cdot83$, and from Table 2 we get $1\cdot18$. Hence

$$P_r \left\{ 0\cdot83 < \frac{\lambda_1 s_1^2 + \lambda_2 s_2^2}{\sigma_1^2 + \sigma_2^2} < 1\cdot18 \right\} \simeq 0\cdot90,$$

or
$$P_r \left\{ \frac{107\cdot61}{1\cdot18} < \sigma_1^2 + \sigma_2^2 < \frac{107\cdot61}{0\cdot83} \right\} \simeq 0\cdot90,$$

or
$$P_r \{ 91 < \sigma_1^2 + \sigma_2^2 < 130 \} \simeq 0\cdot90.$$

Thus 90 % confidence limits for $\sqrt{(\sigma_1^2 + \sigma_2^2)}$ are approximately $9\cdot5$ and $11\cdot4$.

In a similar manner 98 % confidence limits can be found by using Tables 3 and 4.

7. Discussion

In order to use the tables given in this paper in experiments of the type described above, it is necessary that both rows of the analysis of variance table should have degrees of freedom as large as 16. This can be ensured by making the number of levels of the factor large enough. If the number of levels is small we shall not get a good estimate of the variability between levels, and hence we shall be led to wide confidence limits for the gross variability. When the number of levels is very small, they will be so wide that it will not be worth spending time estimating them. Thus it is desirable that the number of levels should be as large as possible.

The ideal experiment for this type of analysis will have two replications for each level and a large number of levels, so that both lines of the analysis of variance table are of comparable accuracy. However, if the cost of increasing the number of replications is negligible, there is no harm done if more than two are made. It will be advantageous to make the number of replications reasonably large, if an accurate estimate of the interaction variance is also required from the analysis.

Although the number of levels, K, should be large, experiments are often made in which K is small. For example, in chemical and engineering tests, K has an average of about 10, rarely rising above 20. There are two reasons for this use of a small number of levels of the factor:

(i) The cost of increasing K is usually greater than the cost of increasing the number of replications. An example of this is found in agricultural field trials where the levels of a factor correspond to farms. Here the cost of transporting personnel and equipment to an extra farm will be greater than the cost of taking an extra observation on the farms which have already been chosen for the experiment.

(ii) There may be more trouble involved in increasing K. For example, a manufacturer may send samples of his product to each of K laboratories, n tests being made on the sample in each laboratory. It will be much easier to carry out a few extra tests than to get a larger number of laboratories to co-operate.

However, there is a way in which large values of the degrees of freedom can arise. The experiment can be repeated several times and the results of the separate experiments combined. This, in fact, was done in the example given above. I think that this technique could profitably be applied to other experiments in which high accuracy is required.

8. CONCLUSION

A series expansion suitable for estimating confidence limits for a general linear combination of variances has been derived as far as the term of order f^{-1}. In numerical work for the case of a positive combination of two variances it has shown itself to be remarkably good, the best results being obtained at the upper end of the distribution. In view of this fact, tables of the 1, 5, 95 and 99 % points have been calculated and are given in this paper. They are suitable for estimating confidence intervals for the sum of two variances, if the degrees of freedom are large enough. An example of their use is given in the text.

REFERENCES

BARTLETT, M. S. (1953). Approximate confidence intervals. II. More than one unknown parameter. *Biometrika*, **40**, 306–17.

CORNISH, E. A. & FISHER, R. A. (1937). Moments and cumulants in the specification of distributions. *Extrait de la Revue de l'Institute International de Statistique*, **4**, 1–14. Reprinted as paper 30 in *Contributions to Mathematical Statistics*, by R. A. Fisher (1950). John Wiley and Sons, Inc.

TIPPETT, L. H. C. (1952). *Technological Applications in Statistics*. Williams and Norgate Ltd.

WELCH, B. L. (1956). On linear combinations of several variances. *J. Amer. Statist. Ass.* **51**, Part 1.

INTERPOLATIONS AND APPROXIMATIONS RELATED TO THE NORMAL RANGE

By JOHN W. TUKEY*

Princeton University

Many quantities related to the normal range are only well known for a rather open grid of values, and interpolation is often considered difficult (vide David, Hartley & Pearson's (1954) remark at p. 486). Methods of convenient interpolation should thus be of use.

Elfving (1947) established a general asymptotic result for ranges of samples from symmetrical distributions. For the normal distribution this implies that

$$n \int_{\frac{1}{2}w}^{\infty} e^{-\frac{1}{2}x^2} dx$$

has a limiting distribution. The usual limiting value of Mill's ratio yields

$$\int_{u}^{\infty} e^{-\frac{1}{2}x^2} dx \sim u^{-1} e^{-\frac{1}{2}u^2}.$$

When we combine these and take logarithms, we learn that

$$w^2 + 8 \ln w - 8 \ln n$$

has a limiting distribution. We might then expect, when the middle term could be neglected, that w^2 would behave like $8 \ln n = 18.42 \log_{10} n$, while the reciprocal of the variance of w would be proportional to

$$\left(\frac{d(w^2 + 8 \ln w)}{dw} \right)^2 = 4w^2 \left(1 + \frac{4}{w^2} \right)^2 \sim 18.42 \log_{10} n + 8.$$

Thus the use of

$$f_{a,b,c}(n) = a \log_{10} b(n+c)$$

would seem to offer promise as a basis for interpolation.

With suitable choices of b and c, the ratios of suitable powers of:

(1) $w_{5\%}$ and $w_{1\%}$, the upper 5 and 1 % points of the normal range,
(2) d_n and V_n, the average value and variance of the normal range,
(3) the ratio $d_n / \sqrt{V_n}$, and
(4) $(w/s)_{5\%}$ and $(w/s)_{1\%}$, the upper 5 and 1 % points of the ratio of a range to s from the same sample,

to $\log_{10} b(n+c)$ are remarkably constant. This is shown in Table 1, where the values of $n > 20$ are those most likely to be key values in view of the existence of standard deviations for the normal range for these values. The ratios have been carried to enough decimal places to show either trends in the ratio, or irregularities which reflect limitations in the available tables.

The constancy of the ratios is so great as to provide not only interpolations but simple and rather accurate approximations. With the growing use of internally programmed automatic

computers the already substantial convenience of such simple analytical approximations is being substantially enhanced.

In view of the surprisingly good approximations obtained by choosing b and c properly, it is important to emphasize that useful interpolation is also quite possible and even simpler, since division by the square root of $\log_{10} n$ will make almost all of the quantities treated easily interpolable for n between 20 and 1000.

The 1% points of the range for $n \geqslant 20$ were taken from Table 22 of the new *Biometrika* tables (based in part on Pearson (1942a), while the corresponding 5% points were interpolated in Table 23 to give a third decimal, values for $30 \leqslant n \leqslant 100$ were taken from Pearson (1932), while values for $n = 200, 500$ and 1000 were computed from the four moments found by Tippett (1925), using Table 42 of the *Biometrika* tables (based in part on Pearson & Merrington (1951)). The values so found were:

n	100	200	500	1000
$w_{5\%}$	6·09	6·49	7·00	7·37
$w_{1\%}$	6·62	7·02	7·50	7·84

(the values for $n = 100$ being included for comparison with Pearson's values of 6·08 and 6·63). The approximations (corresponding to 'std' in Table 1)

$$w_{5\%} = \{17 \cdot 0 \log_{10}(1 \cdot 5n)\}^{\frac{1}{2}},$$
$$w_{1\%} = \{17 \cdot 5 \log_{10}(3 \cdot 3n)\}^{\frac{1}{2}},$$

seem to be about as accurate as the tabulated values for $20 \leqslant n \leqslant 1000$ and to be satisfactory for $n \geqslant 6$. For some purposes, it is more convenient to write the approximations in the form (corresponding to 'alt' in Table 1)

$$(w_n)_{5\%}/(w_2)_{5\%} = 1 \cdot 49\{\log_{10}(1 \cdot 5n)\}^{\frac{1}{2}},$$
$$(w_n)_{1\%}/(w_2)_{1\%} = 1 \cdot 51\{\log_{10}(3 \cdot 3n)\}^{\frac{1}{2}}.$$

The average values of the normal range, d_n, were taken from Table 27 of the *Biometrika* tables (based on Tippett (1925)). The approximation ('std' in Table 1)

$$d_n = \{17 \cdot 06 \log_{10} 0 \cdot 291(n + 2 \cdot 6)\}^{\frac{1}{2}}$$

is good to about 1 part in 1500 for $20 \leqslant n \leqslant 1000$, while interpolation in Table 1 for a more precise ratio than 17·06 will easily give an accuracy of 1 in 20,000. The alternate approximation, which in this case agrees with the approximation just given to 4 significant figures, is

$$d_n/d_2 = 3 \cdot 66\{\log_{10} 0 \cdot 291(n + 2 \cdot 6)\}^{\frac{1}{2}}.$$

The variances of the normal range, V_n, were taken from Table 20 of the *Biometrika* tables (based on Hartley & Pearson (1951)) for $n \leqslant 20$, from Tippett (1925), as repeated on p. 45 of the *Biometrika* tables, for $n = 60, 100, 200, 500$ and 1000, and from Pearson (1932) for $n = 30, 45$ and 75. The approximation

$$1/V_n = 1 \cdot 338 \log_{10} 1 \cdot 05(n + 4 \cdot 4)$$

is apparently about as good as the tabulated values for $20 \leqslant n \leqslant 1000$ and is quite useful for $n \geqslant 6$. The alternate form differs by about 2 parts in 1300 and is

$$V_n/V_2 = 1 \cdot 03/\log_{10} 1 \cdot 05(n + 4 \cdot 4).$$

Table 1. *Ratios of various quantities to* $\log_{10}(b(n+c))$ *for selected values of* n

Quantity	$[w_5,]^2$	$[w_1,]^2$	d_n^2	$d_n \sqrt{V_n}$	$d_n \sqrt{V_n}$	$1 V_n$	$[(w s)_5]^2$	$[(w s)_1]^2$
b	1·5	3·3	0·291	0·52	0·494	1·05	0·78	1·1
c	0·0	0·0	2·6	2·5	$1 \cdot 535 \sqrt[4]{n}$	4·4	−4·5	−7·4
$n = 2$	16·10	16·17	10·0	3·58	6·84	1·66	Imag.	Imag.
3	16·82	17·05	13·5	4·18	4·900	1·423	Imag.	Imag.
4	16·9	17·28	15·0	4·42	4·8340	1·366	Imag.	Imag.
5	17·000*	17·38	15·69	4·56	4·8347	1·3440	Imag.	Imag.
6	17·035*	17·47	16·12	4·630	4·8349	1·3343	133·	Imag.
8	17·037*	17·52	16·57	4·711	4·8351	1·3348	39·	86·
10	17·026*	17·53	16·78	4·7497	4·8351	1·3340	21·	32·
15	17·041*	17·53	16·99	4·7871	4·8353	1·3359	19·04	21·3
20	17·006*	17·54	17·054	4·7984	4·8352	1·3370	18·62	20·10
30	17·06	17·50	17·082	4·8034	4·8345	1·3378	(18·41)	(19·73)
45	16·96	17·47	17·079	4·8029	4·8341	1·3382	n a	n a
60	16·98	17·50	17·070	4·8015	4·8343	1·3382	18·49	19·95
75	17·03	17·49	17·062	4·8001	4·8336	1·3377	n/a	n/a
100	16·99†	17·45‡	17·052	4·8008	4·8355	1·3393	18·59	20·14
200	17·00	17·43	17·036	4·7979	4·8345	1·3388	18·64	20·17
500	17·04	17·48	17·044	4·7951	4·8317	1·3370	18·62	20·14
1000	17·10	17·49	17·056	4·8008	4·8358	1·3396	18·59	20·03
std	17·0	17·5	17·06	4·8	4·835	1·3380	18·6	20·1
alt	17·5	17·52	17·06	—	—	1·336	18·49	20·25

* Based on interpolation in the cumulative distribution to a 3rd decimal.
† 17·04 by method used for $n > 100$.
‡ 17·40 by method used for $n > 100$.
Imag. = imaginary, n/a = no percentage point tabulated.
Values in ( ) based on interpolations by Pearson (1932).

Two sorts of approximations to the reciprocal of the coefficient of variation of the range are given. The first yields

$$d_n/\sqrt{V_n} = 4 \cdot 8 \log_{10} 0 \cdot 52(n + 2 \cdot 5),$$

and is accurate to the accuracy of the known values or to 1 part in 1000 for $n \geqslant 20$. The second, which illustrates an altered form for c, yields

$$d_n/\sqrt{V_n} = 4 \cdot 835 \log_{10} 0 \cdot 494(n + 1 \cdot 535 \sqrt[4]{n}),$$

which is accurate to the accuracy of the known values or to 1 part in 4000 for $n \geqslant 4$. (No alternatives are provided.)

The percentage points of the ratio of a range to the standard-deviation estimate, s, from the same sample were taken from the recent paper by David *et al.* (1954). The approximations

$$(w/s)_{5\%} = \{18 \cdot 6 \log_{10} 0 \cdot 78(n-4 \cdot 5)\}^{\frac{1}{2}},$$

$$(w/s)_{1\%} = \{20 \cdot 1 \log_{10} 1 \cdot 1(n-7 \cdot 4)\}^{\frac{1}{2}},$$

are not quite as good as the key values, the ratios in Table 1 showing a tendency to run low near $n = 50$, and, probably, high near $n = 300$. The amount of this effect is near the limit of accuracy, and should not amount to more than about 1 part in 400 for $20 \leqslant n \leqslant 1000$. About as much more error is introduced by two-significant-figure constants in the alternate forms

$$(w/s)_{5\%} = 4 \cdot 3\{\log_{10} 0 \cdot 78(n-4 \cdot 5)\}^{\frac{1}{2}},$$

$$(w/s)_{1\%} = 4 \cdot 5\{\log_{10} 1 \cdot 1(n-7 \cdot 4)\}^{\frac{1}{2}}.$$

METHODS

The approximations given are not best in any specific sense, but have been chosen to fit reasonably well. Except for $w_{5\%}$ and $w_{1\%}$, where $c = 0$ works quite well, the values of b and c were usually determined to make the ratios for $n = 20$, 100 and 1000 nearly the same, which can be easily done by trial and error solution of

$$\log_{10}(100+c) - \frac{\log_{10}(1000+c) - \log_{10}(100+c)}{(u_{1000}/u_{100}) - 1}$$

$$= \log_{10}(20+c) - \frac{\log_{10}(1000+c) - \log_{10}(20+c)}{(u_{1000}/u_{20}) - 1},$$

the common value of these differences being taken as $-\log_{10} b$.

ILLUSTRATION OF APPROXIMATION

To illustrate the use of the technique in more refined approximation, Fig. 1 shows values of

$$\text{antilog}_{10}(d_n/4 \cdot 835 \sqrt{V_n}) - 0 \cdot 496n - 0 \cdot 6 \log_{10} n$$

plotted against $\log_{10} n$. For larger values of n the dashed vertical lines indicate the uncertainties associated with the fact that V_n is given to only 3 decimals for $n = 60$, 100, 200, 500 and 1000, and to only 4 decimals for 30, 45 and 75. Interpolation to nearly 4-decimal accuracy in V_n is clearly possible from $n = 20$ nearly to $n = 100$. Better values for n's between 100 and 1000 would clearly be obtained from one or two 4 or 5 decimal values of V_n, rather than by many more values of lower accuracy. A single highly accurate quadrature for an n of, say, 1001 would probably go far to settle this situation.

The quantity plotted can be regarded as

$$0 \cdot 496c' - 0 \cdot 6 \log_{10} n,$$

where

$$d_n/\sqrt{V_n} = 4 \cdot 835 \log_{10} 0 \cdot 496(n+c').$$

Here $4 \cdot 835$ was chosen from the second approximation to $d_n/\sqrt{V_n}$, while $0 \cdot 496$ was modified from $0 \cdot 494$ to give a relatively straight plot for n between 10 and 100. The use of $0 \cdot 6 \log_{10} n$ is not essential, and serves merely to simplify the plot somewhat.

A further use of the same technique is in supplementing the available values of the lower percentage points of w s. The lower 2·5% point serves as a good example. Fitting to $n = 10$, 100 and 1000 yields

$$3\cdot907 - \{\log_{10}[(n+11\cdot42)\,7\cdot79]\}^{\frac{1}{4}}$$

as a first approximation. Keeping the 3·907, and adjusting the 7·79 to 8·14 to produce reasonable linearity against $\log n$, leads to the use of

$$z = 8\cdot14\,\text{antilog}_{10}\,(y/3\cdot907)^2 - n - 4\cdot6\log_{10}n,$$

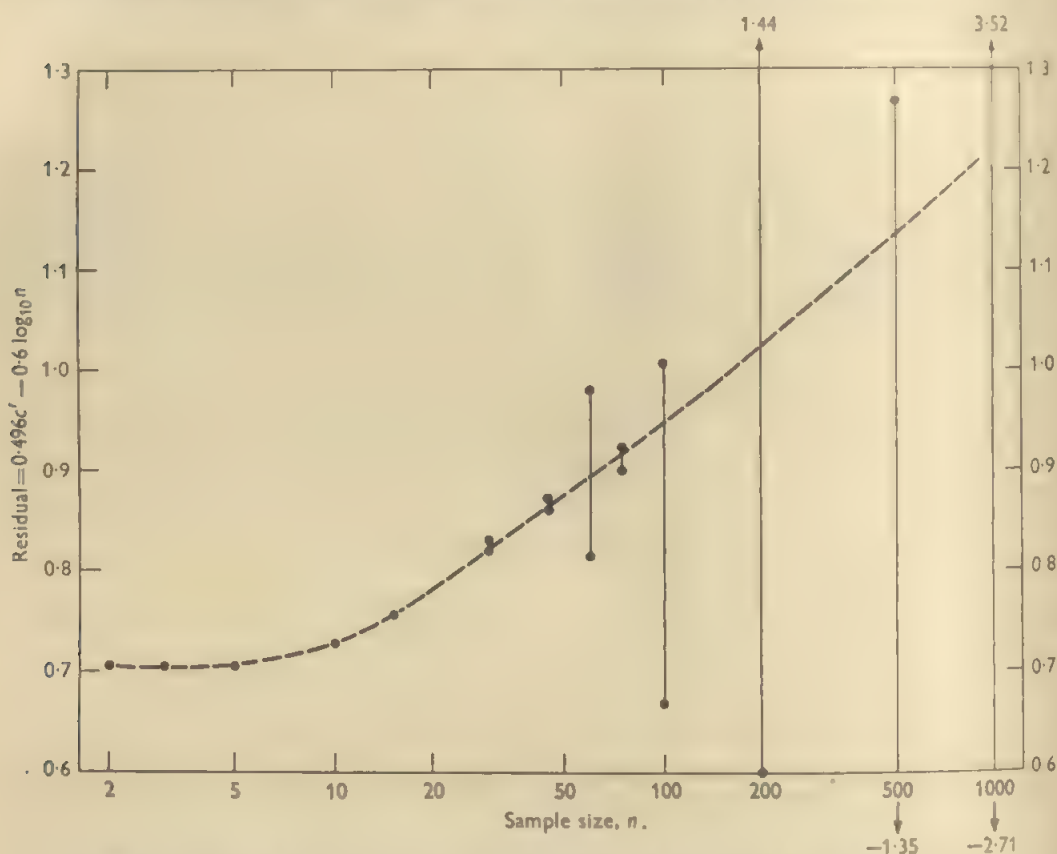as a quantity of reasonable constancy, where y is the lower 2·5% point of w s.

Fig. 1. Residuals for $d_n\sqrt{V_n}$, where c' is defined by $d_n\,V_n = 4\cdot835\,\log_{10}0\cdot496\,(n+c')$.

The numerical values found are as follows:

n	2	3	10	15	20	60	100	200	500	1000
z	7·63	7·69	7·79	8·00	8·70	7·97	7·28	7·87	7·89	43·42

The values for $n = 2$ and 3 are taken from Thomson (1955). There appears to be some ground for suspicion that the values of y for 20 and 1000 are 0·02 or 0·03 high. Interpolation for $n = 4$ to 9 should be simple and effective.

While the original approximation (to $w_{5\%}$) was obtained entirely empirically, the writer owes the suggestion of its asymptotic explanation to his former colleague, David L. Wallace.

REFERENCES

DAVID, H. A., HARTLEY, H. O. & PEARSON, E. S. (1954). *Biometrika*, **41**, 482–93.

ELFVING, G. (1947). *Biometrika*, **34**, 111–19.

HARTLEY, H. O. & PEARSON, E. S. (1951). *Biometrika*, **38**, 463.

PEARSON, E. S. (1932). *Biometrika*, **24**, 404–17.

PEARSON, E. S. (1942). *Biometrika*, **32**, 301–8.

PEARSON, E. S. & MERRINGTON, M. (1951). *Biometrika*, **38**, 4–10.

THOMSON, G. W. (1955). *Biometrika*, **42**, 268–9.

TIPPETT, L. H. C. (1925). *Biometrika*, **17**, 364–87.

THE GAMBLER'S RUIN PROBLEM WITH CORRELATION

By C. MOHAN

London School of Economics and Columbia University, New York

1. Introduction

In the classical problem of the gambler's ruin there is a constant probability of the gambler winning any game and the games are independent. Feller, for example, has treated this problem in his book (1950). We consider here a more general problem where there is a correlation between the results of two successive games. We consider two players X, Y. Conditional on winning the previous game, the probabilities that X wins or loses the next game are p_1, q_1, and conditional on losing the previous game these probabilities are q_2, p_2. Thus X's ability to win or lose is governed by the transition probability matrix (t.p.m.):

$$\begin{array}{cc} & \begin{array}{cc} \text{win} & \text{loss} \end{array} \\ \begin{array}{c} \text{win} \\ \text{loss} \end{array} & \begin{pmatrix} p_1 & q_1 \\ q_2 & p_2 \end{pmatrix}, \end{array} \tag{1.1}$$

where

$$p_1 + q_1 = 1 = p_2 + q_2. \tag{1.2}$$

If we suppose that the stake in each game is one coin, two successive games can have for X one of the following pairs of results, the transition probabilities being given before the colon:

$$p_1: 1 \to 1, \quad q_1: 1 \to -1, \quad p_2: -1 \to -1, \quad q_2: -1 \to 1.$$

On the assumption that a score of $+1$ has the same probability as a score of -1 in the first game, the coefficient of correlation is

$$R = \frac{p_1 - q_2}{\sqrt{(p_1 + q_2)(p_2 + q_1)}}. \tag{1.3}$$

When $p_1 = p_2 = p$ (so that also $q_1 = q_2 = 1 - p = q$), the correlation becomes $R = p - q$. This corresponds to a real situation, as, for example, a game between two players, of whom the one holding the 'bank' has an advantage, the bank passing to the winner of each game. We suppose that the probability of winning with the bank is p, and $p > q$, so that $R > 0$. On the other hand, if the rule were that the bank passes to the loser of each game, we should have a negative correlation R.

Suppose that X starts with a capital r and Y with a capital $\alpha - r$, and the games are played with a certain (positive or negative) correlation R, until one or other of the players is ruined. How is the probability of ruin affected by R? In particular, if X had the choice of determining whether the games were to be played with a positive or negative correlation, on what basis should he decide so as to maximize his chances of winning?

This problem is analogous to a correlated random walk in the presence of absorbing barriers at $n = 0$ and at $n = \alpha$, the particle starting from a position $r (0 < r < \alpha)$ and moving a unit step to the right or to the left according to the t.p.m. (1.1).

The random walk is said to be symmetric if $p_2 = p_1 = p$. Then at the end of each step there is a probability p of a step being in the same direction as the preceding one and a

probability q ($= 1 - p$) of the direction being reversed. We shall consider first the problem for the general unsymmetric case corresponding to the t.p.m. (1·1).

In case the correlation is zero (i.e. $p_1 = q_2 = p$, so that $q_1 = p_2 = 1 - p$), the steps are to the right with a constant probability p and to the left with a constant probability $1 - p$; thus there is a drift towards the right which is measured by $2p - 1$. When $p = \frac{1}{2}$ the random walk is both symmetric and uncorrelated.

In the case of unsymmetrical correlated walk, the drift may still be measured by the quantity

$$\delta = p_1 - p_2.$$

Introducing the quantity

$$c = p_1 + p_2 - 1, \tag{1·4}$$

the coefficient of correlation R given by (1·3) is now given in terms of c and δ by the relation

$$R^2(1 - \delta^2) = c^2. \tag{1·5}$$

It will be noticed that in the symmetrical case ($\delta = 0$) $c = R$.

2. Probability of ultimate ruin

Denote by a_r and b_r the probabilities of ultimate ruin of X (whose initial capital is r) conditional on the first game being a win or a loss. It is easy to see that a_r and b_r satisfy the recurrence relations

$$\left. \begin{array}{l} a_r = p_1 a_{r+1} + q_1 b_{r+1} \\ b_{r+1} = p_2 b_r + q_2 a_r \end{array} \right\} \quad (0 < r < \alpha - 1). \tag{2·1}$$

and

Denoting by E the operation of increasing the suffix r by unity, these may be written as

$$(1 - p_1 E) a_r = q_1 E b_r,$$
$$(E - p_2) b_r = q_2 a_r,$$

which show that a_r and b_r satisfy the difference equation

$$(E - p_2)(1 - p_1 E) \chi_r = q_1 q_2 E \chi_r,$$

i.e.

$$\left\{ E^2 - \frac{1+c}{p_1} E + \frac{p_2}{p_1} \right\} \chi_r = 0.$$

Hence

$$b_r = G + H \left(\frac{p_2}{p_1} \right)^r,$$

$$a_r = \frac{1}{q_2} \left[G + H \left(\frac{p_2}{p_1} \right)^{r+1} - p_2 \left\{ G + H \left(\frac{p_2}{p_1} \right)^r \right\} \right].$$

The boundary values are $b_1 = 1$ and $a_{\alpha-1} = 0$; whence

$$G + H \left(\frac{p_2}{p_1} \right) = 1, \quad q_2 G + q_1 \left(\frac{p_2}{p_1} \right)^\alpha H = 0,$$

so that

$$p_2^\alpha q_1 / G = -q_2 p_1^\alpha / H = p_2(p_2^{\alpha-1} q_1 - q_2 p_1^{\alpha-1}).$$

Substituting these values of G and H we are led to the probabilities of ultimate ruin

$$\left. \begin{array}{l} b_r = \dfrac{q_1 \lambda^\alpha - q_2 \lambda^r}{q_1 \lambda^\alpha - q_2 \lambda} \\[2mm] a_r = \dfrac{q_1 \lambda^\alpha - q_1 \lambda^{r+1}}{q_1 \lambda^\alpha - q_2 \lambda}, \end{array} \right\} \tag{2·2}$$

and

where $\lambda = p_2 / p_1$.

Now if c_1 denotes the probability of initial win and c_2 $(= 1 - c_1)$ of initial loss, then the general expression for P_r, the probability of ultimate ruin unconditional on the result of the first game, is

$$P_r = \frac{c_1 q_1 (\lambda^{\alpha-1} - \lambda^r) + c_2 (q_1 \lambda^{\alpha-1} - q_2 \lambda^{r-1})}{(q_1 \lambda^{\alpha-1} - q_2)}. \qquad (2\cdot3)$$

For $\lambda = 1$, $(2\cdot3)$ is indeterminate. In this symmetrical case the equations $(2\cdot1)$ reduce to

$$a_r = p a_{r+1} + q b_{r+1}$$

and

$$b_{r+1} = p b_r + q a_r.$$

Hence, it is found that P_r is now given by

$$P_r = \frac{c_2 + (\alpha - 1 - r) q}{1 + q(\alpha - 2)}. \qquad (2\cdot4)$$

P_r can be expressed in terms of δ and c, since

$$\lambda = (c + 1 - \delta)/(c + 1 + \delta) \quad \text{and} \quad q_2/q_1 = (1 - c + \delta)/(1 - c - \delta).$$

As the formula in its general form is not easy to interpret, we take for illustration the symmetrical case

$$p_2 = p_1 = p \quad \text{with} \quad c_1 = c_2 = \tfrac{1}{2}, \qquad (2\cdot5)$$

and study the behaviour of P_r as the correlation R varies from -1 to $+1$. Now

$$R = p - q = 1 - 2q = c.$$

Let $\rho = 2R/(1-R)$, so that $q = \tfrac{1}{2}(1-R) = 1/(\rho+2)$. Then from $(2\cdot4)$

$$P_r = \frac{1}{2} \left(1 + \frac{1 - 2\dfrac{r}{\alpha}}{1 + \dfrac{\rho}{\alpha}} \right). \qquad (2\cdot6)$$

Now ρ varies between -1 and $+\infty$ as R varies between -1 to $+1$, and $\alpha > r > 0$. Hence

$$P_r \text{ lies between } \tfrac{1}{2} \text{ and } 1 - (r - \tfrac{1}{2})/(\alpha - 1).$$

Thus $P_r \gtrless \tfrac{1}{2}$ according as $r \lessgtr \tfrac{1}{2}\alpha$. It will be noticed that while $\lim_{R\to 1} P_r = \tfrac{1}{2}$, $\lim_{R\to -0} P_r$ is different in the three cases $r \lessgtr \tfrac{1}{2}\alpha$. The situation is shown below in the graphs of equation $(2\cdot6)$ for three typical cases. Each curve cuts off from the vertical axis an intercept $(1 - r/\alpha)$; as $R \to +1$, $P_r \to \tfrac{1}{2}$; when $R \to -1$, $P_r \to 1 - (r - \tfrac{1}{2})/(\alpha - 1)$. These results are easy to interpret, for when $R \to 1$ the probability of ultimate ruin depends less and less on the initial capital and should more and more approach the probability of losing the very first game; this is shown by the three curves approaching one another as $R \to 1$. The value of P_r for $R = -1$ is actually non-existent, as is borne out by a study of the equations $(2\cdot1)$ which for $R = -1$ degenerate into a single set of equations

$$a_r = b_{r+1} \quad (1 \leqslant r \leqslant \alpha - 2),$$

with the boundary values $a_{\alpha-1} = 0$, $b_1 = 1$. These $(\alpha - 2)$ equations are not soluble; in fact, these show that $a_1, a_2 \ldots a_{\alpha-2}$ are respectively equal to $b_2, b_3, \ldots, b_{\alpha-1}$ and none of them involves either of the unknown probabilities $a_{\alpha-1}, b_1$. On the other hand, when $R = 1$, equations $(2\cdot1)$ become

$$a_r = a_{r+1}, \quad b_r = b_{r+1} \quad (1 \leqslant r \leqslant \alpha - 2),$$

with the boundary values $b_1 = 1$, $a_{s-1} = 0$. Hence

$$a_1 = a_2 = \ldots = a_{s-1} = 0 \quad \text{and} \quad b_1 = b_2 = \ldots = b_{s-1} = 1,$$

so that $\qquad\qquad P_r = c_1 a_r + c_2 b_r = c_2 \quad (= \tfrac{1}{2} \text{ for the graph}).$

We conclude that if X's capital is initially less than Y's, it is advantageous for him to choose the positively correlated play, and if initially greater than Y's, then to choose negatively correlated play. In each case the greater the correlation in magnitude the more it is to be preferred.
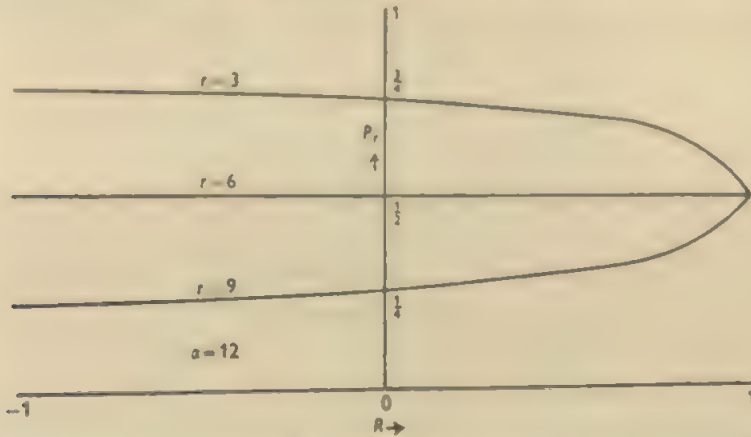

Fig. 1. The graphs show the probability of ruin, P_r, as a function of the correlation R in a game in which the total capital is 12 for three cases of initial capitals of $r = 3$, 6 and 9.

3. EXPECTED DURATION OF PLAY

The expected value of the duration of play can also be determined by the use of difference equations. The method is justified if we assume that these expectations are finite. Let A_r, B_r be the respective durations of play if the initial capital is r and the first game results in a win or loss.

Then if the first game results in a win the capital becomes $r + 1$ and then the expected duration is A_{r+1} with probability p_1 and B_{r+1} with probability q_1. Hence

$$\left.\begin{aligned} A_r &= p_1 A_{r-1} + q_1 B_{r-1} + 1 \\ B_{r+1} &= p_2 B_r + q_2 A_r + 1 \end{aligned}\right\} \quad (0 < r < x - 1). \qquad (3\cdot1)$$

similarly

The boundary values are

$$B_1 = 1 = A_{s-1}. \qquad (3\cdot2)$$

Equations $(3\cdot1)$ become homogeneous on putting

$$A_r = Kr + L + d_r, \quad B_r = Mr + N + f_r, \qquad (3\cdot3)$$

if

$$Kr + L = p_1(Kr + K + L) + q_1(Mr + M + N),$$

$$Mr + M + N = p_2(Mr + N) + q_2(Kr + L) + 1;$$

i.e. if

$$K = p_1 K + q_1 M, \qquad L = p_1(K + L) + q_1(M + N) + 1,$$

$$M = p_2 M + q_2 K, \quad M + N = p_2 N + q_2 L + 1.$$

Remembering the relation (1·2), viz. $p_1 + q_1 = 1 = p_2 + q_2$, these equations are solved by

$$K = M, \quad q_1(L-N) = M+1 \quad \text{and} \quad q_2(L-N) = M-1,$$

i.e.

$$K = M = (q_1 + q_2)(q_1 - q_2) \quad \text{and} \quad L - N = 2(q_1 - q_2). \tag{3·4}$$

This solution shows that the homogeneity of (3·1) achieved by the substitution (3·3) is not disturbed if the same constant is added to both the relations (3·3). This conclusion is corroborated by a study of the structure of equations (3·1): for if A_r and B_r are increased by k (a constant with regard to r) then by virtue of the relation (1·2) k cancels out from them. Consequently one of the constants L and N can be chosen arbitrarily. We therefore make the substitution

$$A_r = Kr + d_r, \quad B_r = Mr + N + f_r,$$

i.e. set $L = 0$ in (3·3). This substitution transforms (3·1) into

$$d_r = p_1 d_{r+1} + q_1 f_{r+1}, \quad f_{r+1} = p_2 f_r + q_2 d_r.$$

Proceeding as in §2 the solution of these equations is found to be

$$d_r = A + B\left(\frac{p_2}{p_1}\right)^r,$$

$$f_{r+1} = \frac{1}{q_1}\left[A + B\left(\frac{p_2}{p_1}\right)^r - p_1\left\{A + B\left(\frac{p_2}{p_1}\right)^{r+1}\right\}\right],$$

so that

$$A_r = \frac{q_1 + q_2}{q_1 - q_2}r + A + B\left(\frac{p_2}{p_1}\right)^r,$$

and

$$B_{r+1} = \frac{q_1 + q_2}{q_1 - q_2}(r+1) - \frac{2}{q_1 - q_2} + A + B\left(\frac{p_2}{p_1}\right)^r \frac{q_2}{q_1}.$$

The boundary values (3·2) give

$$\frac{q_1 + q_2}{q_1 - q_2}(\alpha - 1) + A + B\left(\frac{p^2}{p_1}\right)^{\alpha-1} = 1 = \frac{-2 + q_1 + q_2}{q_1 - q_2} + A + B\frac{q_2}{q_1},$$

whence

$$B = \frac{(\alpha - 2)(q_1 + q_2) + 2}{(q_2/q_1 - \lambda^{\alpha-1})(q_1 - q_2)}, \quad A = \frac{-2q_2 + 2}{(q_1 - q_2)} - \left(\frac{q_2}{q_1}\right)B.$$

Thus

$$A_r = \frac{(q_1 + q_2)r + 2p_2}{q_1 - q_2} + \frac{(\alpha - 2)(q_1 + q_2) + 2}{(q_1 - q_2)}\frac{\lambda^r - q_2/q_1}{q_2/q_1 - \lambda^{\alpha-1}},$$

and

$$B_r = \frac{(q_1 + q_2)r - 2q_2}{q_1 - q_2} + \frac{(\alpha - 2)(q_1 + q_2) + 2}{(q_1 - q_2)}\frac{q_2}{q_1}\frac{\lambda^r - 1}{q_1 q_2/q_1 - \lambda^{\alpha-1}}. \tag{3·5}$$

We now consider the symmetrical case $q_1 = q_2$. The relations (3·4) are indeterminate in this case. We therefore try the quadratic expressions

$$A_r = K_1 + L_1 r + M_1 r^2, \quad B_r = K_2 + L_2 r + M_2 r^2.$$

Substituting these expressions in

$$A_r = pA_{r+1} + qB_{r+1} + 1, \quad B_{r+1} = pB_r + qA_r + 1, \tag{3·6}$$

and identifying the various powers of r,

$$K_1 = p(K_1 + L_1 + M_1) + q(K_2 + L_2 + M_2) + 1,$$

$$L_1 = p(L_1 + 2M_1) + q(L_2 + 2M_2), \quad M_1 = pM_1 + qM_2.$$

$$K_2 + L_2 + M_2 = qK_1 + pK_2 + 1, \quad L_2 + 2M_2 = qL_1 + pL_2, \quad M_2 = qM_1 + pM_2.$$

The third and sixth give $M_1 = M_2$ and the other four are

$$q(K_2 - K_1) + pL_1 + qL_2 + M_1 + 1 = 0, \tag{3.7}$$

$$q(L_2 - L_1) + 2M_1 = 0, \tag{3.8}$$

$$q(K_1 - K_2) - L_2 - M_1 + 1 = 0, \tag{3.9}$$

and one identical with (3.8).

Adding (3.7) and (3.9)

$$p(L_1 - L_2) + 2 = 0, \tag{3.10}$$

so that by (3.8)

$$M_1 = \tfrac{1}{2}q(L_1 - L_2) = -q/p,$$

and finally by (3.9)

$$q(K_1 - K_2) = L_2 + M_1 - 1 = L_2 + 1/p \quad \text{by (3.10)}.$$

Thus

$$\left. \begin{aligned} A_r &= K_1 + r\{q(K_1 - K_2) - 1/p\} - \frac{q}{p}r^2, \\[2mm] B_r &= K_2 + r\{q(K_1 - K_2) + 1/p\} - \frac{q}{p}r^2. \end{aligned} \right\} \tag{3.11}$$

The initial conditions (3.2) give

$$K_1 + (\alpha - 1)\{q(K_1 - K_2) - 1/p\} - (q/p)(\alpha - 1)^2 - 1 = K_2 + q(K_1 - K_2) + 1,$$

which are solved by

Thus (3.11) become

$$K_1 = \alpha \quad \text{and} \quad K_2 = -q\alpha/p.$$

$$\left. \begin{aligned} A_r &= \alpha + (r/p)(\alpha q - 1) - r^2 q/p, \\[1mm] B_r &= -(\alpha + r^2)q/p + (\alpha q + 1)r/p. \end{aligned} \right\} \tag{3.12}$$

and

Hence the expected duration of play unconditional upon the result of the first game is

$$D_r = c_1 \left\{ (\alpha - r) + \frac{q}{p}r(\alpha - 1 - r) \right\} + c_2 \left\{ r + \frac{q}{p}(r - 1)(\alpha - r) \right\}, \tag{3.13}$$

where c_1, c_2 are, as already defined, the initial probabilities of win and loss. Assuming the first win and loss to be equally probable (3.13) becomes

$$D_r = r(\alpha - r) - \frac{R}{1+R}\{-\alpha + 2r(\alpha - r)\}. \tag{3.14}$$

When $R = 0$, $D_r = r(\alpha - r)$, in agreement with Feller's formula (1950, p. 287). Further, since $r(\alpha - r) \geqslant \tfrac{1}{4}\alpha^2 \geqslant \tfrac{1}{2}\alpha$ (as $\alpha \geqslant 2$), it follows that except in the trivial case $\tfrac{1}{2}\alpha = 1 = r$, for positively correlated play the duration has smaller expectation than for the uncorrelated case and for negatively correlated play it has larger expectation. In fact as $R \to 1$, $D_r \to \tfrac{1}{2}\alpha$ and as $R \to -1$, $D_r \to \infty$.

4. Generating functions for the probabilities of the gambler's ruin at the nth game

Denote by $v_{r,n}$ the conditional probability that X (whose initial capital is r) is ruined at the nth game, given that he wins the first game and by $w_{r,n}$ the corresponding probability when he loses the first game.

Evidently

and

$$\left. \begin{aligned} v_{r,n} &= 0 \quad \text{if} \quad r > n-2 \quad \text{or if} \quad n-r \text{ is odd,} \\ w_{r,n} &= 0 \quad \text{if} \quad r > n \quad\quad \text{or if} \quad n-r \text{ is odd.} \end{aligned} \right\} \tag{4.1}$$

Define the generating functions (g.f.) of $v_{r,n}$, $w_{r,n}$

$$V_r(s) = \sum_{n=1}^{\infty} v_{r,n} s^n, \quad W_r(s) = \sum_{n=1}^{\infty} w_{r,n} s^n. \tag{4.2}$$

Elementary reasoning shows that the following recurrence relations hold

$$\left. \begin{array}{l} v_{r,n+1} = p_1 v_{r-1,n} + q_1 w_{r+1,n} \\ w_{r+1,n+1} = p_2 w_{r,n} + q_2 v_{r,n} \end{array} \right\} \quad (1 < r < \alpha - 1,\ n \geqslant 1). \tag{4.3}$$

Evidently

$$w_{1,1} = 1, \quad v_{\alpha-1,n} = 0, \quad n \geqslant 1. \tag{4.4}$$

Define the boundary value

$$w_{1,n} = 0 \quad (n > 1) \tag{4.5}$$

in order that (4.3) may hold for $r = 1$ also. Now multiply the equations (4.3) by s^{n+1} and add from $n = 1$ to ∞, then since $v_{r,1} = 0 = w_{r+1,1}$, $r \geqslant 1$, we have

$$V_r(s) = p_1 s V_{r+1}(s) + q_1 s W_{r+1}(s), \tag{4.6}$$

and

$$W_{r+1}(s) = p_2 s W_r(s) + q_2 s V_r(s).$$

Using the operator E already defined these may be written as

$$(1 - p_1 s E) V_r(s) = q_1 s E W_r(s),$$

$$(E - p_2 s) W_r(s) = q_2 s V_r(s),$$

which show that $V_r(s)$ and $W_r(s)$ satisfy the difference equation

$$(E - p_2 s)(1 - p_1 s E) \chi_r = q_1 q_2 s^2 E \chi_r,$$

i.e.

$$\left\{ E^2 - \frac{1 + cs^2}{p_1 s} E + \frac{p_2}{p_1} \right\} \chi_r = 0.$$

Hence

$$W_r(s) = A \eta_1^r + B \eta_2^r, \tag{4.7}$$

so that by (4.6)

$$V_r(s) = (q_1 s)^{-1} \{ A \eta_1^{r+1} + B \eta_2^{r+1} - p_2 s (A \eta_1^r + B \eta_2^r) \},$$

where η_1, η_2 are the roots of the quadratic

$$p_1 \eta^2 - \frac{(1 + cs^2)}{s} \eta + p_2 = 0.$$

Since (4.6) are homogeneous in $V_r(s)$, $W_r(s)$ and hold for $r = 1, 2, \ldots, \alpha - 2$, they can determine $V_r(s)$ and $w_r(s)$ for $r = 1, 2, \ldots, \alpha - 1$ if two of these $2(\alpha - 1)$ functions are known. These two known functions are $V_{\alpha-1}(s)$ and $W_1(s)$, since on using (4.2) the relations (4.4) and (4.5) lead to the boundary conditions

$$V_{\alpha-1}(s) = \sum_{n=1}^{\infty} v_{\alpha-1,n} s^n = 0, \quad W_1(s) = \sum_{n=1}^{\infty} w_{1,n} s^n = s. \tag{4.8}$$

For these values (4.7) give

$$A \eta_1 + B \eta_2 = s, \quad A \eta_1^{\alpha-1}(\eta_1 - p_2 s) + B \eta_2^{\alpha-1}(\eta_2 - p_2 s) = 0,$$

whence, writing

$$D(s) = D = \eta_1 \eta_2^{\alpha-1}(\eta_2 - p_2 s) - \eta_1^{\alpha-1} \eta_2(\eta_1 - p_2 s),$$

we have

$$DA = s \eta_2^{\alpha-1}(\eta_2 - p_2 s) \quad \text{and} \quad DB = -s \eta_1^{\alpha-1}(\eta_1 - p_2 s).$$

Substituting these values in (4·7) and noting that

$$(\eta_1 - p_2 s)(\eta_2 - p_2 s) = \frac{p_2}{p_1} - \frac{p_2}{p_1}\{1 - s^2(p_1 + p_2 - 1)\} + p_2^2 s^2 = \frac{p_2 q_1 q_2}{p_1} s^2,$$

we have finally

$$V_r(s) = q_1 s^2 D^{-1}(\eta_2^2 \eta_1^{r-1} - \eta_1^2 \eta_2^{r-1}),$$

and

$$W_r(s) = s D^{-1}\{\eta_2^2 \eta_1^r(\eta_2 - p_2 s) - \eta_1^2 \eta_2^r(\eta_1 - p_2 s)\}. \qquad (4·9)$$

It may be noted that the probabilities a_r, b_r of ultimate ruin calculated in the second section may be determined from these functions $V_r(s)$ and $W_r(s)$. In fact,

$$a_r = \sum_{n=r+2}^{\infty} v_{r,n} = \sum_{n=1}^{\infty} v_{r,n} = V_r(1),$$

and

$$b_r = \sum_{n=1}^{\infty} w_{r,n} = W_r(1).$$

Similarly, we have

$$A_r = V_r'(1) \quad \text{and} \quad B_r = W_r'(1).$$

5. OBSERVATIONS ON THE CASE $\alpha = \infty$

(i) *Probability of ultimate ruin*

Evidently when $p_2 > p_1$, i.e. when the drift δ is negative, as $\alpha \to \infty$, a_r and b_r as given in (2·2) tend to unity. Even when $p_2 = p_1$, i.e. the probabilities of a win or loss being continued are equal, the formula (2·2) show that with probability unity, X will be ruined; but when $p_2 < p_1$, then the probability of ultimate ruin, according as the first game is won or lost is

$$\lambda^r \frac{q_1}{q_2} \quad \text{or} \quad \lambda^{r-1}.$$

(ii) *Duration of play*

(a) When $p_2 > p_1$, i.e. when the drift δ is negative, as $\alpha \to \infty$, the last terms in (3·5) tend to zero and

$$A_r \to \frac{(q_1 + q_2)r + 2p_2}{q_1 - q_2} \quad \text{and} \quad B_r \to \frac{(q_1 + q_2)r - 2q_2}{q_1 - q_2}.$$

(b) When $p_2 < p_1$, i.e. when there is a positive drift δ, as $\alpha \to \infty$, (3·5) show that A_r and $B_r \to \infty$.

(c) When $p_2 = p_1$, as $\alpha \to \infty$, (3·12) shows that A_r and $B_r \to \infty$ and hence from (3·13) $D_r \to \infty$, unless $c_1 = 0 = q$. This exceptional case occurs when with probability unity the first game is lost, as also is each succeeding game, so that X is ruined in r games.

In conclusion, I wish to express my gratitude to Dr F. G. Foster, who suggested this problem to me and also supervised the investigation.

REFERENCE

FELLER, W. (1950). *An Introduction to Probability Theory and its Applications*. New York: Wiley and Sons.

TABLES FOR SIGNIFICANCE TESTS OF 2×2 CONTINGENCY TABLES

By P. ARMSEN

National Institute for Personnel Research, South African Council for Scientific and Industrial Research

1. INTRODUCTION

In a 2×2 tables of the type

	Y	Not Y	
X	a	c	A
Not X	b	d	B
	R	S	N

it is assumed that a, b, c and d are the absolute frequencies resulting from a double dichotomy of N individuals according to the properties X and Y.

The null hypothesis to be tested is:

There is no association between X and Y. If this is true, one would expect in the long run of experiments with the same marginal totals A, B, R, S, to have $a : \bar{b} = A : B$. or which is the same, $N . \bar{a} = A . R$.

The whole 2×2 table can—for fixed marginal totals—be determined by one of the entries. It is arbitrary which one would choose. The probability of the set of frequencies observed and of possible other sets of frequencies which might have been observed can be computed by the exact hypergeometric formula given by Fisher (1941, p. 95).* From this it follows that the possible sets of frequencies can be looked at as events with a completely known one-variate discrete probability distribution. It is bell-shaped or—in some cases — J-shaped, e.g.
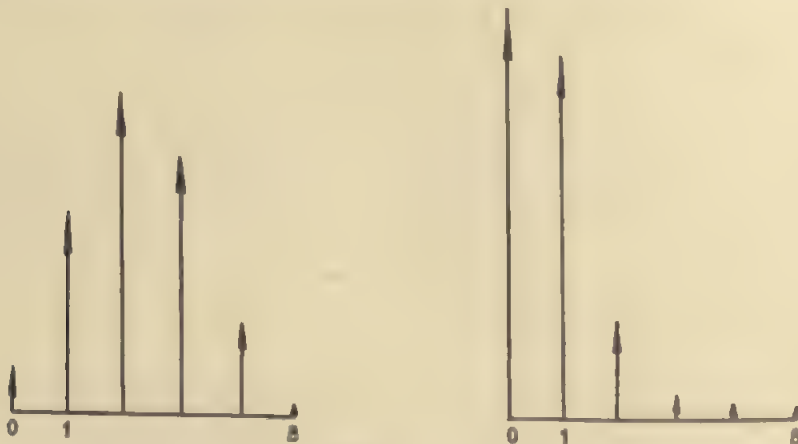

Figure 1.

* There are detailed discussions of the applicability of the formula in Barnard (1947) and Cochran (1952).

The range of the distribution of b or d is from 0 to B if B is the smallest marginal total. The present tables are based on direct evaluation of the hypergeometric formula.

It is possible to look at the given null hypothesis in two ways:

(1) The object is to determine experimentally whether or not there is reasonable evidence for a positive association between X and Y (or, which is the same, for a negative association between X and Not Y). In this case it is natural and generally agreed to take the observed table as significant (for departure from the null hypothesis) at, say, the 100α % level, if the sum of the probabilities of the observed and the possible more extreme tables (extreme in the same direction) is smaller than or equal to α, where, for example, $\alpha = 0.05$ or 0.01. This is called 'the case of one-tailed significance'.

(2) There is no *a priori* knowledge about the relationship between X and Y. Here a decision must be made before starting the experiment as to which of the possible events (sets of frequencies of 2×2 tables) are to be taken as evidence for an existing association between X and Y (positive or negative whichever the case may be).

In this case we are looking out for alternatives to the null hypothesis which may throw the value of a frequency in the 2×2 table, say d, into either of the tails of the probability distribution. The usual procedure is to cut from each tail a number of terms for which the probability on the null hypothesis sums to $\leqslant \frac{1}{2}\alpha$. If the observed value falls in either of these tails we say that the result is significant at the 100α % level, using a 'two-tailed' test.

In analogous problems where the probability distribution is continuous it is possible to find significance points, whether one-tailed or two-tailed, defining rejection regions associated precisely with any desired value of α. For discontinuous variation which arises when dealing with the hypergeometric type of distribution of our present problem or with the binomial or Poisson distribution, this is not the case.

The position is illustrated in Example 1 given in the following section. If we take $\alpha = 0.05$, and look for significance in the sense of a positive relation between X and Y, we shall only establish this if $d = i$ is 6 or 7. But the probability of this result is only $p(6) + p(7) = 0.71$ %, far below the 5 % aimed at. If we use a two-tailed test based on cutting off regions from the two ends of the distribution for each of which the sum of the probabilities is $\leqslant 2.5$ %, then we see that this region will again only include $i = 6$ and 7.

Unless we adopt a procedure of random sampling for a decision, discussed, for example, by Tocher (1950) and Cochran (1952), there is no means of obtaining exactly 5 % for the rejection region. It will, however, be noted that while for the single-tailed test the region is clearly defined, for the two-tailed test alternative regions are possible if we drop the usual convention of making the probability associated with each tail region sum to $\frac{1}{2}\alpha$ or less. Thus coming back to Example 1, it might be asked whether there is serious objection to including $i = 0$, 6 and 7 in the two-tailed 5 % rejection region, giving a total probability of $2.58 + 0.71 = 3.29$ % (still under 5 %) on the null hypothesis?

The object of this paper is to consider alternative definitions of the rejection regions for the two-tailed test, and also to provide tables of 5 and 1 % significance levels for both one- and two-tailed tests, and for 2×2 tables containing up to $N = 50$ observations. These tables, by adopting a new method of entry, make it possible to go beyond the range of existing tables using relatively little space.

2. Choice of the rejection region for the two-tailed test of significance

The definition of a two-tailed 100α % significance region referred to in the introductory section, which has been used largely, e.g. in Finney's (1948) tables, may be expressed as follows:

Definition D_1. Cut as much as possible but not more than $\frac{1}{2}\alpha \times 100$ % from each tail of the distribution under the null hypothesis.

While in any particular case, when the hypergeometric probabilities have been calculated, it may seem clear on intuitive grounds how to enlarge this rejection region by including one or more further values of d and yet keep the total probability below 100α %, when preparing tables for general use it is necessary to have some clearly defined, consistent rule of procedure.

A possible procedure which the author prefers is given by the following definition which is that used in the mathematical description of the tables (pp. 502–5 below):

Definition D_2. Arrange the possible events (defined say in terms of d) in ascending order of the size of their probabilities under the null hypothesis; include in the 100α %, two-tailed rejection region those events for which the cumulative sum of these ordered probabilities is smaller than or equal to α. (The kind of association—positive or negative—which is likely to exist will be indicated by the observed event.)

An equivalent visual description, which can be illustrated on the diagrams given above, is as follows:

Lift a parallel to the horizontal axis until the sum of the probability ordinates falling below this line has reached as closely as possible, but not exceeded α. Each event whose associated probability ordinate is now completely below the parallel is two-tailed significant at the 100α % level. The others are not.

The advantages of this definition D_2 are that:

(i) It is straightforward to apply.

(ii) It will often include more points in the rejection region than will D_1, without increasing the overall probability above α.

(iii) It is consistent in the sense that if an observed event is significant at the $100\alpha_1$ % level, then it must also be significant at any $100\alpha_2$ % level, where $\alpha_2 > \alpha_1$.

(iv) There is a simple way of generalization to the case of multivariate discontinuous distributions, e.g. for the $h \times k$ contingency table.

It can, however, lead to certain anomalies which, as Prof. E. S. Pearson has pointed out to me, may not be regarded as acceptable. Consider the three examples given below.

The definition D_3 will be given later. In Example 1, D_2 includes $i = 0$ in the rejection region which D_1 does not, and this seems very reasonable. For the two-tailed 5 % level in Example 2, D_2 again gives a larger region than D_1, although it is open to question whether it would not have been better to include $i = 0$ rather than $i = 5$ in the region. When, however, we take a 7·2 % level* in this example, we find a more serious difficulty. Although

$$p(0) = 3{\cdot}53 \% < \tfrac{1}{2} \times 7{\cdot}2 \%,$$

D_2 does not make $i = 0$ significant, but chooses $i = 5$ instead. Thus in this instance D_2 contains no term or terms from the lower tail of the distribution.

* It should be made clear that anomalies of this kind arise only rarely and therefore we have illustrated this situation on Example 2, taking an unconventional significance level.

Example 1

a	c	A			33
b	d	B		i	7

$$\frac{}{R \quad S \quad | \quad N} = \frac{}{25 \quad 15 \quad | \quad 40}$$

If d is replaced by a variate i and the marginal totals are kept constant, the probability distribution of i (under the null hypothesis) is as follows:

i	0	1	2	3	4	5	6	7
$p(i)$ in %	2·58	14·25	29·92	30·87	16·84	4·83	0·67	0·04

Definition	Cases of i which are two-tailed significant at the 5 % level
D_1	6, 7
D_2	0 6, 7
D_3	0 6, 7

Example 2

		33
	i	7
26	14	40

i	0	1	2	3	4	5	6	7
$p(i)$ in %	3·53	17·29	32·10	29·19	13·96	3·49	0·42	0·02

Definition	Significant events (values of i)	
	Two-tailed 5 % level	Two-tailed 7·2 % level
D_1	6, 7	0, 6, 7
D_2	5, 6, 7	5, 6, 7
D_3	0, 6, 7	0, 6, 7

Example 3

$$
\begin{array}{c|c}
 & 31 \\
i & 12 \\
\hline
\end{array}
$$

| | 31 | 12 | 43 |

i	0	1	2	3	4	5	6	7	8	9	10	11	12
$p(i)$ in %	0·920	6·624			...		4·435	0·877	0·102	0·006	0	0	0

Definition	Two-tailed significant at the 1% level
D_1	8 9 10 11 12
D_2	7 8 9 10 11 12
D_3	8 9 10 11 12

If we consider what is the objection to the D_2 region in this last respect, we seem to reach the following broad conclusions. The purpose of a two-tailed test is to pick out departures from the null hypothesis in either direction, i.e. to establish significance when there is either a marked positive or negative association between X and Y. Clearly it can only act in this way if the rejection region contains terms from both tails of the distribution. In the case of extreme skewness where the first (or last) term has a probability that is greater than 100α %, 'two-tailedness' is obviously impossible, but in cases like those of Examples 2 and 3 it is *possible* to have a two-tailed 5 % (Ex. 2) and 1 % (Ex. 3) rejection region. What is less clear is how to define that region in general terms.

It is possible that a more fundamental attack on the problem could be based on a study of the power function of the test.* We shall, however, be content here to put forward a definition giving a region which must include that resulting from the customary D_1 but which, as with D_2, will often contain more points. If we start from the two-tailed region of D_1 there will often be only one choice of a point to be included in the wider region: this is the case for Example 1, where the 5 % rejection region under D_1 (including $i = 6, 7$) can only be extended by including $i = 0$. In the case of Example 2, however, the 5 % region of D_1 (6, 7) can be extended either by including $i = 0$ or 5 but not, of course, both.

A number of alternative definitions were considered; some of these were ambiguous or could lead to the inconsistency referred to under (iii) on p. 496 above. Finally, we have taken a definition D_3, which appears generally, though not quite always, to satisfy intuitive requirements for a two-tailed test.

Definition D_3. Define $F(E)$, or the 'first tail' probability, as the cumulative sum of the probabilities under the null hypothesis of all possible events which are more extreme, in the

---

* [In this connexion it is perhaps relevant to note that when considering, on the basis of the power function, an 'unbiased' two-tailed test in the case of continuous asymmetrical variation. Neyman & Pearson (*Statist. Res. Mem.* **2** (1936), 18–25) found that a larger probability (i.e. $> \frac{1}{2}\alpha$) should be cut off from the steeper tail of the null hypothesis distribution than from the finer tail.—ED.]

same direction, than a given event E, including the probability of E itself. Define $S(E)$, or the 'second tail' probability, as the cumulative sum of probabilities starting with that of the event most extreme in the opposite direction as compared with E, and cumulating up to but not exceeding the value of $F(E)$. If, and only if, $F(E) + S(E) \leq x$, include E in the rejection region for the two-tailed $100x\%$ level of significance.

The rejection regions defined by D_3 are shown below the table of probabilities for each of the Examples 1–3.

It can be readily seen that every event which is significant under D_1 will be significant under D_3, but the regions under D_2 and D_3 will not necessarily correspond. This last point is illustrated in Example 2, where D_3 may be thought to meet the requirements for a two-tailed best better than D_2. On the other hand, in the case of Example 3, where a two-tailed 1% rejection region is required, D_3 gives a region no wider than D_1, failing to include the point $i = 7$. It is interesting to note that in this case there is a genuinely two-tailed region which could be used, namely, that including $i = 0, 9, 10, 11, 12$, giving a total probability of $0.926 < 1.000\%$. From the point of view of all-round power of discrimination it is possible that this last region might have some claim for consideration.

To illustrate anomalies and differences between definitions, we have naturally picked out exceptional cases, but it should be emphasized that these cases are very rare. Thus, among all the 2×2 tables with $N \leq 50$ considered in building up the tables of significance points given below, there were only thirty-two cases in which the regions defined by D_2 and D_3 differed (one case of disagreement usually results in two changes in the tables).

In forming the tables with significance levels of 5 and 1%, the computation was done independently three times, once on the basis of D_2 and later on the basis of D_3 and of another alternative definition, not given here, which was finally discarded for possible inconsistency. The tables contain the results according to D_3. The cases where the result of using D_2 is different are indicated by $+$ or $-$ signs above or below the corresponding entry; thus a $+$ indicates that the entry is to be increased by 1, a $-$ that it is to be decreased by 1, to obtain the limit under D_2.

In the following section we give directions for use of the tables with an example, while an Appendix contains the mathematical description of the method by which the tables with this special form of entry have been built up.

3. DIRECTIONS FOR USE OF THE TABLES

$N \leq 50$ is the number of observations.

The tables are sometimes applicable in cases $N > 50$. See Note 2.

No 2×2 table where $N < 6$ is significant one-tailed at the 5 % level.

Step 1. Arrange the given 2×2 table in the form

a	c		A
b	d		B
R	S		N

where A is the largest and B is the smallest (or in case of equality one of the smallest) of the four marginal totals, A, B, R, S, and $aB > Ab$.

If $aB = Ab$, the table is not significant.

Step 2. Note the four figures:

$$d \qquad \text{for finding the appropriate table,}$$

$$\left. \begin{array}{l} b \\ c - b = S - B = A - R \geqslant 0 \end{array} \right\} \quad \text{for finding the place in the table,}$$

$$a - d = R - B = A - S \geqslant 0 \quad \text{for decision.}$$

Step 3. Look at the table at the end of this paper headed $i = d$ and find the intersection of arrays $x = b$ and $y = c - b$. There are four entries

$$z_1 \quad z_2 \quad \text{(exceptions see Notes 1, 2 and 3)}$$
$$z_3 \quad z_4$$

for the one-tailed and two-tailed points of significance according to the scheme:

	One-tailed	Two-tailed
5 % 1 %	z_1 z_3	z_2 z_4

Step 4. Compare the figure $a - d$ with the entry z_j, in which you are interested ($j = 1, 2, 3, 4$), e.g. the given 2×2 table is two-tailed significant at the 5 % level, if and only if $a - d \geqslant z_2$. Correspondingly for z_1, z_3 and z_4.

Note 1. If no entry is given, the answer is:

(*a*) Not significant for $N \leqslant 50$ if the missing entry should be to the right or below any given entry in the table headed $i = d$, or somewhere to the right of the whole table; e.g. $i = 11, x = 5, y = 14, a - d = 2$. The missing entry should be to the right of the given entry $z_1 = 8, i = 11, x = 3, y = 14$. Therefore the event $\begin{vmatrix} 13 & 19 \\ 5 & 11 \end{vmatrix}$ is not significant at the 5 % level.

(*b*) Significant two-tailed at the 1 % level if the missing entry should be to the left or above any given entry in the table headed $i = d$ or somewhere to the left of the whole table; e.g. $i = 16, x = 2, y = 13, a - d = 1$. Because $x < 3$, the missing entry should be to the left of the whole table for $i = 16$. Therefore the event $\begin{vmatrix} 17 & 15 \\ 2 & 16 \end{vmatrix}$ is significant two-tailed at the top level.

(*c*) The entry '–' instead of a numerical entry stands for 'not significant for $N \leqslant 50$'.

Note 2. These tables are sometimes applicable in cases where $B + S \leqslant 50 < N$. i.e. in cases where an entry is given or missing to the left. These tables are not applicable if $50 < B + S$. In most practical cases (assuming $N > 50$) where these tables fail to give an answer, the Table VIII given by Fisher & Yates (1949) gives an answer and vice versa in all cases for $N \leqslant 50$.

Note 3. In cases where only two entries (counting '–' as an entry) are given instead of four, each entry stands for both the one-tailed and two-tailed point of significance.

Example for application of the tables

Given

4	18		22
12	9		21

16	27		43

Step 1

18	9		27
4	12		16

22	21		43

$B = 16 < 21 < 22 < 27,$

$A = 27 > 22 > 21 > 16,$

$a = 18, \quad b = 4,$

$Ba = 16 \times 18 = 288 > 108 = 27 \times 4 = Ab;$

consequently the arrangement is correct.

The only possible other arrangement after having chosen B as the smallest marginal total is

9	18		27
12	4		16

21	22		43

But now $a = 9, \quad b = 12,$

$Ba = 16 \times 9 = 144 < 324 = 27 \times 12 = Ab,$

contrary to the agreed kind of arrangement

Step 2

$$d = 12 \qquad \text{for finding the appropriate table,}$$
$$\left. \begin{array}{l} b = x = 4 \\ c - b = y = 5 \end{array} \right\} \quad \text{for finding the place in the table,}$$
$$a - d = 6 \qquad \text{for comparison with } z_j.$$

Step 3. In the table headed $i = 12$ at the intersection of $x = 4$ and $y = 5$, the entries

$$\begin{array}{cc} z_1 & z_2 \\ z_3 & z_4 \end{array} \quad \text{are given as} \quad \begin{array}{cc} 1 & 2 \\ 6 & 8 \end{array}$$

Step 4. Result: the given 2×2 table is

significant (one-tailed) at the 5 % level	$(6 > 1)$,
significant (two-tailed) at the 5 % level	$(6 > 2)$,
significant (one-tailed) at the 1 % level	$(6 = 6)$,
not significant (two-tailed) at the 1 % level	$(6 < 8)$.

The entry $z_2 = 2$ shows that even the less significant table with the same d, B, S, (b and c), and with $a - d = 2$ or

14	9		23
4	12		16

18	21		39

is significant (two-tailed) at the 5 % level. But $z_1 = 1$ shows that the table with $a - d = 0$ or

12	9		21
4	12		16

16	21		37

is not significant (one-tailed) at the 5 % level.

$z_4 = 8$ shows that the first 2×2 table with the same d, B and S as the given table, which is significant (two-tailed) at the $1\,\%$ level, is with $a - d = 8$ or

20	9	29
4	12	16
24	21	45

Note 4. A user who wants to apply definition D_3 must neglect all $+$ or $-$ signs attached to any entry in the table.

A user who wants to apply definition D_2 must read a $+$ sign attached to an entry as: add 1 to the given entry, and a $-$ sign as: subtract 1 from the given entry. For example:

In the case $i = 5$, $x = 1$, $y = 12$, by $z_2 = \overline{21}$ is meant:

$$\text{in case of } D_3 \quad z_2 = 21$$
$$\text{in case of } D_2 \quad z_2 = 21 - 1 = 20$$

Correspondingly, in the case $i = 9$, $x = 0$, $y = 21$, by $z_4 = \underset{+}{9}$ is meant:

$$\text{in case of } D_3 \quad z_4 = 9$$
$$\text{in case of } D_2 \quad z_4 = 9 + 1 = 10$$

APPENDIX

Mathematical description of the tables

The tables are completely based on the well-known exact formula:

$$p\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \frac{A!\,B!\,R!\,S!}{N!\,a!\,b!\,c!\,d!} = \frac{\dbinom{S}{d}\dbinom{R}{B-d}}{\dbinom{N}{B}} = p(d \mid BSN),$$

where $p(d \mid BSN)$ is the (single) probability of obtaining d with given marginal totals B and S and given number of observations N, assuming that both classifications are independent of each other.

It follows directly from the arrangement of the table [$B = \text{Min}\,(ABRS)$ and $Ba > Ab$] that $Nd > BS$ and $d \leqslant B \leqslant S \leqslant N - B$, in particular $2B \leqslant N$.

For given d, B, S and N the probability $p(i \mid BSN)$ $(i = 0, 1, 2, \ldots, B)$ is a step function of i only. This function, call it $p(i)$, has the following properties:

It rises monotonically from $i = 0$ to a maximum which is at $i_{max.} = [m]$, where

$$(N+2)\,m = (B+1)\,(S+1)$$

and falls monotonically until $i = B$.

There are some special cases:

(1) $i_{max.}$ may be 0 or B, in which cases $p(i)$ is monotonically decreasing or increasing in its whole range.

(2) There are two equal values $p(i_{max.}) = p(i_{max.} - 1)$ in the cases where m is an integer.

Proof.

$$p(i) : p(i+1) = n : D, \text{ where } n = (i+1)\,(N-S-B+i+1) \text{ and } D = (B-i)\,(S-i).$$

Thus

$$p(i) \leqslant p(i+1) \text{ if } v(i) \leqslant 0,$$
$$p(i) \geqslant p(i+1) \text{ if } v(i) \geqslant 0, \quad \text{where} \quad v(i)\,(N+2) = n - D, \ v(i) = i + 1 - m.$$

Therefore,

$$\text{if } i \geqslant i_{max.} \quad v(i) \geqslant [m] + 1 - m > 0,$$
$$\text{if } i < i_{max.} \quad v(i) \leqslant [m] - m \leqslant 0,$$

which proves the above statements.

It may be shown that

$$\left[\frac{BS}{N}\right] < i_{max} \leqslant \left[\frac{BS}{N}\right] + 1.$$

Definitions $f = f(d \mid BSN)$ or   'first tail':

$$f(d \mid BSN) = \sum_{i=d}^{B} p(i \mid BSN)$$

$s = s(d \mid BSN)$ or   'second tail':

$$s(d \mid BSN) = \sum_{i=0}^{k} p(i \mid BSN), \quad \text{where} \quad k < d,$$

$$p(k \mid BSN) \leqslant p(d \mid BSN), \quad p(k+1 \mid BSN) > p(d \mid BSN) \quad (k+1 = d \text{ in the case } d = i_{max}).$$

If no k exists (that is, if $p(0 \mid BSN) > p(d \mid BSN)$), then $s(d \mid BSN) = 0$.
[k is (given B and S) a function $k(d, N)$ and $Nk < BS$.]

$$t = t(d \mid BSN) \quad \text{or} \quad \text{'two-tailed probability'},$$
$$t = f + s.$$

From these definitions follows:

Statement 1. $p(i \mid BSN) > p(i \mid BS, N+1)$ if $i \geqslant d$
Statement 2. $f(d \mid BSN) > f(d \mid BS, N+1)$ } exactly.
Statement 3. $p(i \mid BSN) < p(i \mid BS, N+1)$ if $i \leqslant k(d, N)$
Statement 4. $t(d \mid BSN) \geqslant t(d \mid BS, N+1)$ *not exactly, but with sufficient approximation in practice.*

The construction of the tables is based on Statements 2 and 4.

For the one-tailed case, it would be enough to compute $f(d \mid BSN)$ for fixed d, B, S and increasing N until the table is found to be significant. Then Statement 2 enables the tables to be condensed to the given form. For the two-tailed case, Statement 4 could serve the same purpose if it were exact. The inexactness of Statement 4 was the reason for computing all possible cases until $N = 50$. The computation was done by direct evaluation of the exact formula for $t(d \mid BSN)$.

Thus the entries z_1 and z_3 give answers based on mathematically exact general reasons. The entries z_2 and z_4 are (although of course correct in the tables) based only on empirical computation. An exact proof of Statement 4 is impossible, as one can show examples where it is not true.

Proofs for Statements 1, 2 and 3

(1) $(h + BS) . p(i \mid BSN) = [h + i(N+1)] . p(i \mid BS, N+1)$, where $h = (N+1)(N+1-B-S)$. Statement 1 follows on account of $i(N+1) > iN \geqslant dN > BS$.

(2) Follows directly from 1.

(3) $$\frac{BS}{N} - \frac{BS}{N+1} = \frac{BS}{N(N+1)} \leqslant \frac{B(N-B)}{N(N+1)} \leqslant \frac{N}{4(N+1)} < \frac{1}{4}.$$

Corresponding to 1, one finds:

$$*p(i, N) \geqslant p(i, N+1) \quad \text{as long as} \quad (N+1)i \geqslant BS,$$

and for $$(N+1)i < BS, \quad p(i, N) < p(i, N+1).$$

But for the single possible value i with $\dfrac{BS}{N+1} \leqslant i < \dfrac{BS}{N}$ follows $i = i_{max}$, because

$$m - \frac{BS}{N+1} = \frac{B+S+1}{N+2} - \frac{BS}{(N+1)(N+2)} < 1,$$

and this, together with $i \leqslant k(d, N)$, contradicts the definition of k which proves Statement 3.

Conclusions from Statement 3

(1) $k(d, N)$ is a monotonically decreasing step function of N, or $k(d, N) \geqslant k(d, N+1)$.

(2) As long as $p[k(d, N) \mid BS, N+1] \leqslant p(d \mid BS, N+1)$, it follows that $s(d \mid BSN) < s(d \mid BS, N+1)$. but in the cases where $p[k(d, N) \mid BS, N+1] > p(d \mid BS, N+1)$, there is one member less in $s(N+1)$ than in $s(N)$, and thus $s(d \mid BSN) > s(d \mid BS, N+1)$.

(It is easy to prove that $p(i \mid BSN) > p(i-1 \mid BS, N+1)$ if $i \leqslant k(d, N)$.)

* By $p(i, N)$ is meant $p(i \mid BSN)$.

Thus the second tail as a function of N increases monotonically step by step in intervals of N, where

$$k(d, N+1) = k(d, N),$$

and decreases in big steps from N to $N+1$, where

$$k(d, N+1) < k(d, N)$$

(or where the length of the second tail decreases).

Of course, $\lim\limits_{N \to \infty} s(d \mid BSN) = 0$, because $\lim\limits_{N \to \infty} i_{\max.} = 0$.

Question of Statement 4

$\lim\limits_{N \to \infty} t(d \mid BSN) = 0$, because $t = s + f$, $\lim\limits_{N \to \infty} s = 0$, and $\lim\limits_{N \to \infty} f = 0$, the last equation following from:

$$\lim\limits_{N \to \infty} i_{\max.} = 0 \quad \text{and} \quad \lim\limits_{N \to \infty} p(0, N) = \lim\limits_{N \to \infty} \frac{\binom{N-S}{B}}{\binom{N}{B}} = 1.$$

The question now is: Is this decreasing of $t(d \mid BSN)$ with increasing N monotonic like that of $f(d \mid BSN)$ or not? Unfortunately it is not.

Proof. Computation gives, e.g. for

$$d = 9, \quad B = 9, \quad S = 36, \quad N = 48,$$
$$t(9 \mid 9, 36, 48) < 0 \cdot 0883,$$
$$t(9 \mid 9, 36, 49) > 0 \cdot 0892.$$

Therefore, Statement 4 is certainly not exact.[*]

The size of $t(N)$ and $t(N+1)$ is influenced mainly by the sums

$$\tau_N = p(d \mid BSN) + p(k \mid BSN) \quad \text{and} \quad \tau_{N+1} = p(d \mid BS, N+1) + p(k \mid BS, N+1),$$

where both times $k = k(d, N)$.[†]

The second term in τ_{N+1} is included in $t(N+1)$ only under the condition

$$p(d, N+1) \geqslant p(k, N+1).$$

Thus only these cases must be considered.

We have

$$\tau_N = \frac{\phi_N + \sigma_N}{D_N},$$

where

$$\phi_N = \binom{S}{d} \binom{N-S}{B-d}, \quad \sigma_N = \binom{S}{k} \binom{N-S}{B-k}, \quad D_N = \binom{N}{B},$$

$k = k(d, N)$ in σ_{N+1} also.

It is easy to show by simple computing that

$$\tau_N . D_{N+1} \left(1 - \frac{B}{N+1}\right) = \phi_{N+1} \left(1 - \frac{B-d}{N+1-S}\right) + \sigma_{N+1} \left(1 - \frac{B-k}{N+1-S}\right),$$

and the ratio τ_N / τ_{N+1} will be smallest (that is, the 'bad' case for Statement 4) if

$$\phi_{N+1} = \sigma_{N+1} \quad \text{or} \quad \tau_{N+1} D_{N+1} = 2\phi_{N+1},$$

and in this worst case one has

$$\tau_N \left(1 - \frac{B}{N+1}\right) = \tau_{N+1} \left\{1 - \frac{2B - (d+k)}{2(N+1-S)}\right\}.$$

Thus, in the worst case $\tau_N < \tau_{N+1}$ is equivalent to

$$\frac{B}{N+1} < \frac{2B - (d+k)}{2(N+1-S)} \quad \text{or} \quad 2BS > (N+1)(d+k).$$

[*] This difficulty does not arise if the definition D_2 is replaced by D_1, because then Statement 4 is exact. But in case of D_3, this example applies and D_3 does not improve the situation as compared with D_2.

[†] If k does not exist, then $s = 0$ and Statement 4 is exact.

Now both the members in this 'bad' inequality are usually of roughly the same size, because

$$\text{if} \quad 2S \leqslant N+1 \quad \text{then} \quad d+k \leqslant B,$$
$$\text{if} \quad 2S \geqslant N \quad \text{then} \quad d+k \geqslant B,$$

and in the case $2S = N+1$ both members are equal and therefore

$$t(d \mid B, S, 2S) = t(d \mid B, S, 2S-1),$$

and only in the rare cases where the three conditions:

(1) $\phi_{N+1} \approx \sigma_{N+1}$,
(2) $\phi_{N+1} \geqslant \sigma_{N+1}$,
(3) $2BS > (d+k)(N+1)$,

are all given is it possible to have $\qquad \tau_{N+1} > \tau_N$,

and therefore it is only in these cases to be expected that $t(N+1) > t(N)$.

Empirically, it was found that, for $N \leqslant 50$ and the levels of 5 and 1 %, significance, no inconsistency is introduced to the tables by using Statement 4, in spite of its inexactness.

The author wishes to thank Prof. E. S. Pearson, to whom he owes a redrafting of §§ 1 and 2, Mr A. G. Arbous and Mr J. E. Kerrich, for their critical remarks and helpful suggestions, and also the members of the Mathematical Statistics Department of the National Institute for Personnel research, for their assistance in computational work and editing of this paper.

REFERENCES

BARNARD, G. A. (1947). Significance tests for 2×2 tables. *Biometrika*, **34**, 123–38.
COCHRAN, W. G. (1952). χ^2 test of goodness of fit. *Ann. Math. Statist.* **23**, 315–45.
FINNEY, D. J. (1948). Tests of significance in a 2×2 contingency table. *Biometrika*, **35**, 148.
FISHER, R. A. (1941). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
FISHER, R. A. & YATES, F. (1949). *Statistical Tables for Biological, Agricultural and Medical Research*. Edinburgh: Oliver and Boyd.
LATSCHA, R. (1953). Tests of significance in a 2×2 contingency table: extension of Finney's tables. *Biometrika*, **40**, 74.
MAINLAND, D. (1948). Statistical methods in medical research. Tables V and VI. *Canad. J. Res. E*, **26**, 1–116.
TOCHER, K. D. (1950). Extension of the Neyman-Pearson theory of tests to discontinuous variates. *Biometrika*, **37**, 130–44.

i = 2

X	top (0 1 2)	bottom (0 1 2)
0	3 14 29	11 37 −
1	7 20 39	20 − −
2	10 27	30 −
3	14 33	39 −
4	17 39	48 −
5	21	57
6	24	66
7	28	−
8	31	−
9	35	−

i = 3

X	top (0 1 2 3 4)	bottom (0 1 2 3 4)
0	0 1 6 13 23 36	4 4 14 30 − −
1	2 2 9 18 29	8 8 21 38 −
2	4 4 12 22 34	12 12 27 − −
3	6 6 15 26	15 15 33 −
4	8 8 18 31	19 19 − −
5	9 9 21 35	23 23 − −
6	11 11 24	26 26 −
7	13 13 27	30 30 −
8	15 15 30	34 34 −
9	16 16 33	− − −
10	18 18	− −
11	20 20	− −
12	21 21	− −
13	23 23	− −
14	25 25	− −
15	27 27	− −
16	28 28	− −

i = 4

X	top (0 1 2 3 4 5)	bottom (0 1 2 3 4 5)
0	0 0 2 3 7 13 21 30	1 1 7 7 16 27 − −
1	0 0 4 5 10 17 25	4 4 11 11 21 34 −
2	1 3 6 6 13 20 30	6 6 15 15 26 − −
3	2 4 8 8 15 24	8 8 18 18 31 −
4	4 5 10 10 18 27	10 10 22 22 − −
5	5 6 12 12 21 31	12 12 25 25 − −
6	6 7 14 14 24	15 15 29 29 −
7	7 8 16 16 26	17 17 33 33 −
8	8 9 18 18 29	19 19 − − −
9	9 10 20 20	21 21 − −
10	10 11 22 22	23 23 − −
11	11 12 24 24	26 26 − −
12	13 13 26 26	28 28 − −
13	14 14	− −
14	15 15	− −
15	16 16	− −
16	17 17	− −
17	18 18	− −
18	19 19	− −
19	20 20	− −
20	22 22	− −

i = 1

X	top (0 1)	bottom (0 1)
0	18 76	98 396
1	37 114	197 594
2	56 153	296 −

i = 5

X	top (0 1)	bottom (0 1)
0	0 1	4 4
0	0 0 2 2	1 2 6 6
1	0 0 3 3	3 3 9 9
2	0 1 4 5	4 4 11 11
3	1 1 6 9	6 6 14 14
4	2 2 7 10	7 7 16 16
5	3 3 9 12	9 9 19 19
6	4 4 10 13	10 10 21 21
7	4 5 12 15	12 12 24 24
8	5 6 13 16	13 13 26 26
9	6 7 14 18	15 15 28 28
10	7 8 16 19	17 17 − −
11	8 9 17 21	18 18 − −
12	9 10 19 22	20 20 − −
13	9 11 20 24	21 21 − −
14	10 12 21 −	23 23 − −
15	11 14	24 24
16	12 15	− −
17	13 16	− −
18	13 18	− −
19	14 −	− −
20	15 −	− −
21	16 −	− −
22	17 −	− −

ARRANGEMENT OF THE ENTRIES

	ONE-TAILED	TWO-TAILED
5 %	z_1	z_2
1 %	z_3	z_4

$a - d \geq z_j$, significant

$< z_j$, not significant

D_3: neglect all + or − signs

D_2: read ± as $\left\{ \begin{array}{c} add \\ subtract \end{array} \right\}$ 1

$i = 5$

X\	2	3	4	5	6
0	4 4	8 11	14 16	20 22	28 29
	10 10	17 –	12 21	– –	– –
1	6 8	– 13	17 19	24 25	
	13 13	22 22	– –	– –	
2	8 10	13 16	20 22	28 29	
	17 17	26 26	– –	– –	
3	10 12	16 18	24 26		
	20 20	30 30	– –		
4	12 14	18 20	26 27		
	23 23	– –	– –		
5	14 16	21 23			
	26 26	– –			
6	16 19	23 25			
	30 30	– –			
7	17 20	26 27			
	– –	– –			
8	19 22				
	– –				
9	21 24				
	– –				
10	23 26				
	– –				
11	25 –				
	– –				

$i = 6$

X\	1	2	3	4	5	6
0	0 0	1 3	5 6	9 10	14 16	20 24
	1 2	6 6	12 12	18 18	20 26	– –
1	0 0	3 4	7 7	12 12	21 –	29 –
	3 4	8 8	15 15	22 22	– –	– –
2	1 1	4 5	9 9	14 15	20 21	
	5 6	11 11	18 18	26 26	– –	
3	0 0	2 4	6 7	11 11	16 18	22 –
	2 2	7 7	13 13	21 21	– –	– –
4	0 0	3 5	7 8	13 14	19 21	
	3 5	9 9	16 16	24 24	– –	
5	0 0	4 6	9 9	15 16	21 24	
	4 6	11 11	18 18	27 27	– –	
6	0 1	5 7	10 11	16 18	23 –	
	5 7	13 13	21 21	– –	– –	
7	1 2	6 8	12 13	18 20		
	6 8	14 14	23 23	– –		
8	2 3	7 9	14 14	20 23		
	8 9	16 16	26 26	– –		
9	2 5	9 10	15 16	22		
	9 10	18 18	– –	– –		
10	3 6	10 11	17 18			
	10 11	20 20	– –			
11	4 6	11 12	18 20			
	11 12	22 22	– –			
12	4 7	12 13	20 21			
	12 13	24 24	– –			
13	5 8	13 14	21			
	13 14	– –				
14	6 8	14 15				
	15 15	– –				
15	6 9	15 16				
	16 16	– –				
16	7 10	16 17				
	17 17	– –				
17	8 10	17 18				
	18 18	– –				
18	8 11	18 –				
	19 19	– –				
19	9 12					
	– –					
20	10 12					
	– –					
21	10 13					
	– –					
22	11 14					
	– –					
23	12 14					
	– –					
24	12 –					
	– –					
25	13 –					
	– –					

$i = 7$

X\	1	
0	0 0	
	0 1	
1	0 0	
	1 2	
2	0 0	
	2 3	
3	0 1	
	4 4	

X\	0	3
4	0 0	1 2
	1 2	5 6
5	0 0	2 3
	2 2	7 8
6	0 0	3 3
	3 3	8 11
7	0 0	3 4
	3 4	10 13
8	0 2	4 5
	4 5	11 14
9	1 2	5 6
	5 6	13 15
10	1 3	6 7
	6 7	14 17
11	2 3	7 8
	7 8	16 18
12	2 4	8 9
	8 9	17 19
13	2 4	9 10
	9 10	19 21
14	3 5	10 11
	10 11	20 –
15	4 5	11 13
	11 12	– –
16	4 6	12 14
	12 14	– –
17	5 6	12 15
	13 15	– –
18	5 7	13 –
	14 16	– –
19	6 7	14 –
	15 –	– –
20	6 8	
	16 –	
21	7 8	
	– –	
22	7 9	
	– –	
23	8 9	
	– –	
24	8 10	
	– –	
25	9 10	
	– –	
26	9 –	
	– –	

$i = 7$

X\	2	3	4	5	6	7
0	0 1 / 3 3	2 3 / 8 10	6 7 / 13 15	10 13 / 19 20	15 17 / - -	20 22 / - -
1	1 2 / 5 6	4 4 / 10 12	8 11 / 16 17	12 15 / 23 23	17 20 / - -	-
2	2 3 / 7 10	5 6 / 13 14	10 13 / 19 19	14 17 / - -	20 22 / - -	-
3	3 4 / 9 11	7 9 / 15 16	12 15 / 22 22	17 19 / - -	-	-
4	4 5 / 11 13	9 12 / 17 18	13 16 / - -	19 21 / - -	-	-
5	6 7 / 13 15	10 13 / 20 20	15 18 / - -	21 - / - -	-	-
6	7 8 / 15 17	12 15 / 22 22	17 20 / - -	-	-	-
7	8 10 / 17 18	13 17 / - -	19 22 / - -	-	-	-
8	9 13 / 19 20	15 18 / - -	-	-	-	-
9	11 14 / 21 22	16 20 / - -	-	-	-	-
10	12 16 / - -	18 - / - -	-	-	-	-
11	13 17 / - -	-	-	-	-	-
12	14 18 / - -	-	-	-	-	-
13	16 19 / - -	-	-	-	-	-
14	17 - / - -	-	-	-	-	-

$i = 8$

X\	2	3	4	5	6	7
0	0 0 / 1 1	0 1 / 5 6	3 5 / 9 10	7 8 / 14 16	11 12 / 20 -	15 16 / - -
1	0 0 / 3 4	1 2 / 7 7	5 7 / 12 12	9 10 / 17 20	13 14 / - -	18 19 / - -
2	0 0 / 4 5	3 5 / 9 9	6 8 / 14 15	10 12 / 20 -	15 16 / - -	-
3	0 0 / 2 2	1 1 / 6 7	4 6 / 11 11	8 10 / 16 18	12 13 / - -	17 19 / - -
4	0 0 / 3 5	2 3 / 7 8	6 8 / 13 14	10 11 / 19 -	14 15 / - -	-
5	0 0 / 4 6	3 6 / 9 9	7 9 / 15 16	11 13 / 21 -	16 17 / - -	-
6	0 0 / 0 1	0 1 / 5 7	4 7 / 11 11	8 10 / 17 19	13 14 / - -	18 - / - -
7	0 0 / 1 1	1 2 / 7 8	5 8 / 12 13	9 11 / 19 21	14 15 / - -	-
8	0 0 / 2 2	2 2 / 8 9	6 9 / 14 15	11 12 / - -	16 17 / - -	-
9	0 0 / 3 3	3 3 / 9 10	7 10 / 16 17	12 14 / - -	17 - / - -	-
10	0 0 / 4 4	3 4 / 10 11	8 11 / 17 19	13 15 / - -	-	-
11	0 1 / 4 7	4 6 / 11 12	9 12 / 19 -	15 16 / - -	-	-
12	0 1 / 5 7	5 8 / 13 13	10 13 / - -	16 - / - -	-	-
13	0 2 / 6 8	6 9 / 14 14	11 14 / - -	-	-	-
14	1 2 / 7 9	7 10 / 15 16	12 15 / - -	-	-	-
15	1 2 / 8 10	7 10 / 16 17	13 - / - -	-	-	-
16	2 3 / 8 10	8 11 / 14 -	-	-	-	-
17	2 3 / 9 11	9 12 / - -	-	-	-	-
18	3 4 / 10 12	10 13 / - -	-	-	-	-
19	3 4 / 11 13	10 - / - -	-	-	-	-
20	4 5 / 11 13	11 - / - -	-	-	-	-
21	4 5 / 12 -	-	-	-	-	-
22	4 5 / - -	-	-	-	-	-
23	5 6 / - -	-	-	-	-	-
24	5 6 / - -	-	-	-	-	-
25	6 7 / - -	-	-	-	-	-
26	6 7 / - -	-	-	-	-	-
27	7 - / - -	-	-	-	-	-

$i = 9$

X\	0	1
4	0 0 / 1 2	
5	0 0 / 2 3	
6	0 0 / 3 3	
7	0 0 / 4 4	
8	0 0 / 5 5	
9	0 0 / 1 1	1 1 / 6 6
10	0 0 / 1 3	1 2 / 7 8
11	0 0 / 2 4	2 4 / 8 9
12	0 0 / 3 4	3 5 / 9 10
13	0 0 / 3 5	3 5 / 10 12
14	0 0 / 4 5	4 6 / 11 15
15	0 0 / 5 6	5 7 / 12 -
16	0 1 / 5 6	5 7 / 13 -
17	0 1 / 6 7	6 8 / - -
18	0 1 / 7 8	6 9 / - -
19	1 2 / 7 8	7 9 / - -
20	1 2 / 8 10	8 9 / - -
21	2 2 / 9 -	8 - / - -
22	2 3 / 9 10	
23	2 3 / - -	
24	3 4 / - -	
25	3 4 / - -	
26	4 4 / - -	
27	4 5 / - -	
28	4 - / - -	

$a - d \geq z_j$, significant
     $< z_j$, not significant

D_3: neglect all $+$ or $-$ signs

D_2: read $\pm$ as $\left\{ \begin{array}{c} \text{add} \\ \text{subtract} \end{array} \right\}$!

$i=9$

```
 ⊗ |  3   4    5    6    7    8
 ⊗ 0| 0 0  1   3 4   5   7 8  11 16 16 ·
⊗2 0| 2 3 6 7 10 12 15 16  ·  ·  ·  ·
 1 | 0 0 0 2  4 6 7  9 11 13 17
   | 0 2 4 4  8 10 13 14 18 19 · · · ·
 2 | 0 0 1 3  4 5 7 8 11 13 15 ·
   | 2 3 6 6 10 12 15 16 · · · ·
 3 | 0 0 2 4  5 6 9 10 13 17
   | 3 4 7 10 12 14 17 18 · ·
 4 | 0 1 3 5  6 7 10 12 15
   | 5 5 9 11 14 16 · · · ·
 5 | 1 3 4 6  8 9 12 15
   | 6 7 11 13 16 17 · · · ·
 6 | 2 4 5 7  9 10 14 ·
   | 7 9 13 14 18 · · ·
 7 | 3 4 6 8 11 12 15
   | 9 11 14 16 · · · ·
 8 | 4 5 8 9 12 14
   | 10 13 16 17 · ·
 9 | 4 6 9 10 13 ·
   | 12 14 · · · ·
10 | 5 7 10 11
   | 13 15 · ·
11 | 6 8 11 12
   | 14 16 · ·
12 | 7 9 12 14
   | 16 · · ·
13 | 8 9 13 ·
   | · · · ·
14 | 9 10
   | · ·
15 | 10 11
   | · ·
16 | 11 12
   | · ·
```

$i=10$

```
 ⊗ |  3   4    5    6    7    8
 0 | 0 0 0 1  2 3 5 6 8 11 12 13
   | 0 1 3 5  7 8 12 12 16 · · ·
 1 | 0 0 0 2  3 4 6 9 10 13
   | 2 2 5 7  9 10 14 15 · ·
 2 | 0 0 1 3  5 5 8 11 12 14
   | 3 5 7 8 11 12 16 · · ·
⊗2 3| 0 0 0 1 3 4 6 8 10 12
   | 1 1 5 6  9 9 13 14 · ·
 4 | 0 0 1 2  4 5 7 10 11 14
   | 2 3 6 7 11 11 15 · · ·
⊗1 5| 0 1 2 3 5 6 9 12 13
   | 3 6 8 9 12 13 · · · ·
 6 | 0 0 0 1 3 4 6 8 10 13
   | 1 1 5 7  9 10 14 15 · ·
 7 | 0 0 0 2 4 5 8 11 12 ·
   | 2 2 6 8 11 11 · · ·
 8 | 0 0 1 3 5 6 9 12
   | 2 3 7 9 12 13 · ·
⊗0 9| 0 0 0 2 3 6 7 10 13
   | 3 4 6 10 13 16 · · ·
10 | 0 0 0 0 3 4 7 8 11 ·
   | 0 1 4 6  9 11 · · ·
11 | 0 0 0 2 4 5 8 10
   | 0 1 5 8 11 12 · ·
12 | 0 0 0 2 4 5 9 ·
   | 1 2 6 8 12 13 · ·
13 | 0 0 1 3 5 6 10 ·
   | 1 2 7 9 13 · · ·
14 | 0 0 2 3 6 7
   | 2 3 8 10 · ·
15 | 0 0 2 4 7 8
   | 2 3 9 11 · ·
16 | 0 0 3 5 8 9
   | 3 4 10 12 · ·
17 | 0 0 3 5 8 ·
   | 3 4 11 · · ·
18 | 0 0 4 6
   | 4 5 · ·
19 | 0 0 4 6
   | 5 5 · ·
20 | 0 0 5 7
   | 5 6 · ·
21 | 0 5 6 7
   | 6 6 · ·
22 | 0 1 6 ·
   | 6 7 · ·
23 | 0 1
   | 7 ·
24 | 1 7
   | · ·
25 | 1 2
   | · ·
26 | 1 2
   | · ·
27 | 2 3
   | · ·
28 | 2 ·
   | · ·
```

$i=11$

```
 ⊗ | 2
 2 |
⊗2 3| 0 0
   | 0 1
 4 | 0 0
   | 1 3
 5 | 0 0
   | 1 3
 6 | 0 0
   | 2 4
⊗7 | 0 0
   | 3 5
 8 | 0 0 0
   | 0 1 4 5
 9 | 0 0 0 1
   | 1 2 5 6
10 | 0 0 1 2
   | 2 4 7 7
11 | 0 0 1 2
   | 3 5 8 8
12 | 0 0 2 3
   | 4 5 9 9
13 | 0 1 3 4
⊗0 | 4 6 10 11
14 | 0 0 1 4 4
   | 0 2 5 7 · ·
15 | 0 0 0 2 4 5
   | 0 1 6 7 · ·
16 | 0 0 1 2 5 6
   | 1 1 7 8 · ·
17 | 0 0 1 3 6 7
   | 1 2 8 9 · ·
18 | 0 0 2 3 6 ·
   | 2 2 8 · · ·
19 | 0 0 2 4
   | 2 3 · ·
20 | 0 0 3 4
   | 3 3 · ·
21 | 0 0 3 5
   | 3 4 · ·
22 | 0 0 4 ·
   | 4 5 · ·
23 | 0 0
   | 4 5
24 | 0 0
   | · ·
25 | 0 0
   | · ·
26 | 0 0
   | · ·
27 | 0 1
   | · ·
28 | 0 ·
   | · ·
```

$i = 11$

```
✗│ 4    5    6    7    8
 │0 0  0 1  2 4  5 8  9 11
0│1 3  5 5  9 10 13 - . .
 │0 0  1 2  4 6  7 9  11 -
1│3 4  6 7  11 13 - - - -
 │0 0  0 2  3 5  8 9  11
2│1 3  4 5  8 10 13 - - -
 │0 0  1 1  3 6  7 9  10 -
3│2 4  6 6  10 13 - - - -
 │0 0  2 2  5 7  8 10
4│4 5  8 8  12 14 - -
 │0 1  3 4  6 8  10 11
5│5 6  9 11 - - - -
 │1 1  4 7  7 9
6│6 7  11 13 - -
 │2 2  5 8  8 11
7│8 8  12 - - -
 │2 3  6 9  10 -
8│9 10 - - - -
 │3 5  7 10
9│10 12 - -
  │4 7  8 -
10│11 - - -
  │5 8  9 -
11│- - - -
  │6 9
12│- -
  │7 -
13│- -
  │8 -
14│- -
```

$i = 12$

```
✗│ 4    5    6    7    8
 │0 0  0 0  2 3  5 6  6
0│0 1  2 3  6 8  10 11 - -
 │0 0  0 0  2 4  5 6  8 9
1│1 2  4 5  8 9  - - - -
✗│ 3
 │0 0  0 1  3 5  6 7
2│2 3  6 8  9 11 - -
 │0 0  0 0  1 4  4 6  7 9
3│0 1  4 4  7 9  11 - - -
✗│ 2
 │0 0  0 0  2 5  5 7
4│1 2  5 7  9 10 - -
 │0 0  0 0  1 2  4 6  7 8
5│0 1  3 3  6 8  10 - - -
 │0 0  0 0  2 4  5 7  8 -
6│0 1  4 4  8 10 - - - -
 │0 0  0 0  3 5  6 8
7│1 2  5 6  9 11 - -
 │0 0  0 1  4 6  7 -
8│2 3  6 9  10 - - -
 │0 0  1 3  4 7
9│3 4  7 10 - -
  │0 0  0 0  2 5  5 7
10│0 2  4 5  8 - - -
  │0 0  0 0  3 5  6 -
11│1 2  5 6  - - - -
  │0 0  0 1  4 6
12│1 3  6 7  - -
  │0 0  1 1  4 7
13│2 3  7 9  - -
  │0 0  1 2  5 -
14│3 4  8 - - -
  │0 0  2 3
15│4 4  - -
  │0 0  3 4
16│4 5  - -
  │0 1  3 -
17│5 6  - -
  │0 1  4 -
18│6 6  - -
  │0 0  0 7
19│0 1  0 0
  │0 0  1 2
20│1 1  - -
  │0 0  1 2
21│1 2  - -
  │0 0  2 -
22│2 3  - -
  │0 0
23│2 3
  │0 0
24│- -
  │0 0
25│- -
  │0 0
26│- -
```

$i = 13$

```
✗│ 5
 │0 0
0│0 1
 │0 0
1│2 4
✗│ 4
 │0 0
2│3 5
 │0 0  0 1
3│1 2  5 6
✗│ 3
 │0 0  0 2
4│3 5  6 7
 │0 0  0 1  3
5│1 1  4 6  7 8
 │0 0  0 2  2 4
6│2 2  5 7  - -
 │0 0  0 2  3 5
7│3 5  6 8  - -
✗│ 2
 │0 0  0 1  3 4 6
8│0 1  4 6  8 - - -
 │0 0  0 2  2 4 5 -
9│1 1  5 6  - - - -
  │0 0  0 2  3 5
10│2 2  6 7  - -
  │0 0  1 3  4 -
11│3 4  7 - - -
  │0 0  0 0  1 4
12│0 1  3 6  - -
  │0 0  0 0  2 4
13│0 1  4 7  - -
  │0 0  0 0  3 -
14│1 2  5 - - -
  │0 0  0 1
15│1 2  - -
  │0 0  1 2
16│2 3  - -
  │0 0  1 -
17│3 3  - -
  │0 0  2 -
18│3 4  - -
  │0 0
19│- -
  │0 0
20│- -
  │0 0
21│- -
  │0 0  0 -
22│0 1  - -
  │0 0
23│0 -
  │0 0
24│- -
```

ARRANGEMENT OF THE ENTRIES

	ONE-TAILED	TWO-TAILED
5%	z_1	z_2
1%	z_3	z_4

$a - d \geq z_j$, significant
$ < z_j$, not significant

D_3: neglect all $+$ or $-$ signs

D_2: read $\pm$ as $\left\{ \begin{array}{l} add \\ subtract \end{array} \right\}'$

$i = 13$

✗	6	7	8
0	0 1 / 3	4 5	
	3 5 7 8	- -	
1	0 2 2 4	5 6	
	5 6 9 9	- -	
2	1 3 4 5		
	7 8 - -		
3	2 4 5 6		
	8 9 - -		
4	3 5 6 -		
	- - - -		
5	4 6		
	- -		
6	5 -		
	- -		

✗ 5, $i = 14$

✗	6	7	8	9
0	0 0 / 2	3 4 -		
5	0 / 3 4	5 . .	. .	
1	0 0 0 0	2 3 4		
	0 2 3 4	6 7 -	-	
2	0 0 0 /	3 4 -		
	1 3 4 5	. . -	- -	
4	3	0 0 0 /	3 4	
	2 4 6 6	- -		
4	0 0 0 0	/ 2 4 -		
	0 2 4 5	. . .		
5	0 0 0 /	2 3		
	2 3 5 6	. .		
3	6	0 0 0 2	3 4	
	3 4 6 -	. .		
7	0 0 0 0	/ 3		
	1 2 4 5	- .		
8	0 0 0 /	2 3		
	1 3 5 6	- -		
9	0 0 0 2	3 -		
	2 4 -	. -		
10	0 0 / 2		
	3 5 - .		
2	11	0 0 0 /	2 3
	1 1 4 5	- .	
12	0 0 0 /		
1	3 - -		
13	0 0 0 2		
	2 4 - .		
14	0 0 / -		
	3 - - -		
✗ 1	15	0 0 / -	
	- - - -		
16	0 0 0 /		
	0 1 - -		
17	0 0 0 -		
	1 1 - -		
18	0 0 0 -		
	1 2 - -		
19	0 0		
	- -		
20	0 0		
	- -		

$i = 15$

✗	6	7	8	9
0	0 0 0 0	0 / 2 .		
	0 / 2 3	. . .		
1	0 0 0 0	/ 2		
	1 2 4 5	- .		
2	0 0 1 0 2	. . -	- .
	✗ 5 2	2 3	. . -	. -
3	0 0 0 0	/ 1		
	0 / 3 4	- -		
4	0 0 0 0 2	.		
	1 2	. .		

$i = 16$ (continued) $i = 17$ $i = 18$

Note: If no entry is given, the answer is:

(a) Not significant for $N \leqslant 50$ if the missing entry should be to the right or below any given entry in the table headed $i = d$, or somewhere to the right of that table.

(b) Significant two tailed at the 1% level if the missing entry should be to the left or above any given entry in the table headed $i = d$, or somewhere to the left of that table. For $i \geqslant 18$ all 2×2 tables which are possible with $N \leqslant 50$, are two-tailed significant at the 1% level.

MISCELLANEA

A note on moving ranges

By H. A. DAVID

Commonwealth Scientific and Industrial Research Organization, Sydney,
and Department of Statistics, University of Melbourne

1. Introduction

Consider a production process in which it takes some time to generate a single item for measurement. Under these conditions moving averages and the corresponding moving ranges provide simple current measures of location and dispersion. Moreover, with the customary assumption of an underlying normal distribution, moving average and range charts may be constructed and used in a very similar fashion to ordinary mean and range charts (see, for example, Grant, 1946; Duncan, 1952). One point of difference is that for the preliminary estimation of the population standard deviation it is convenient to use the set of moving ranges, in samples of size n, that become available as the charts are put in operation. To obtain this estimate of σ the average of the set, the *mean moving range* $\omega_1(n)$, need merely be multiplied by the well-known scale factor converting a range into an unbiased estimate of σ. Clearly $\omega_1(2)$ is the mean successive difference. $\omega_1(n)$ will, in general, increase with n in its sensitivity to any trends in the data; but in the absence of such trends $\omega_1(5)$, for example, will be shown to be considerably more efficient than $\omega_1(2)$.

Corresponding to moving ranges we may speak of moving maxima and minima which analogously to charts of ordinary maxima and minima (Howell, 1949) are capable of application to quality control.

This note deals with some properties of moving ranges, maxima and minima. In particular, an expression is obtained for the correlation between two ranges in overlapping samples in terms of means, variances and covariances of order statistics. For a normal parent population this result is applied to the determination of the efficiency of $\omega_1(n)$. The distribution of runs of equal ranges, which tend to occur in moving range charts, is also briefly discussed. It is realized that these devices have not been used widely and this study is offered in the hope that a clearer understanding of the methods and properties may facilitate the decision as to whether their wider application is worth while.

2. Moving maxima

Suppose that $x_{(1)}, x_{(2)}, \ldots, x_{(i)}, \ldots, x_{(N)}$ represent observations arranged in order of time. We may call $(x_{(i)} \ldots x_{(i+n-1)})$ *the ith moving sample of n*, and shall denote the corresponding maximum, minimum and range by $m_i(n)$, $m_i'(n)$ and $w_i(n)$ respectively.

The order statistics in a sample of n will be written, following Godwin (1949), as

$$x(1 \mid n) \geqslant x(2 \mid n) \geqslant \ldots \geqslant x(n \mid n);$$

or more briefly as

$$x_1 \geqslant x_2 \geqslant \ldots \geqslant x_n.$$

To obtain the joint probability

$$P = \Pr[m_i(n) < X, \; m_{i+d}(n) < Y] \quad (0 < d < n)$$

it is only necessary to consider the $(n+d)$ values from which m_i, m_{i+d} are calculated. We find

$$P = F^n(X) F^d(Y) \quad (X \leqslant Y),$$

$$= F^d(X) F^n(Y) \quad (X \geqslant Y),$$

where

$$F(X) = \Pr(x < X).$$

However, we proceed to determine the correlation between m_i and m_{i+d} by a less direct method. The x's will henceforth be assumed continuous.

Consider first two successive maxima and the set $x_1, x_2, \ldots, x_{n+1}$, say, on which they are based. Then $m_1(n) = m_2(n) = x_{[1\,|\,n+1]}$ unless x_1 or x_{n+1} is the largest value in the set. In the latter case, which has probability $2/(n+1)$, m_1, m_2 are $x_{[1\,|\,n+1]}, x_{[2\,|\,n+1]}$. It follows that

$$\mathcal{E}[m_1(n)\,m_2(n)] = \frac{n-1}{n+1}\,\mathcal{E}_{x_{[1\,|\,n+1]}^2} + \frac{2}{n+1}\,\mathcal{E}_{x_{[1\,|\,n+1]}\,x_{[2\,|\,n+1]}}.$$

Writing x_i for $x(i\,|\,n+d)$ we find likewise for the case $d = 2$

$$\mathcal{E}[m_1(n)\,m_3(n)] = \frac{n-2}{n+2}\,\mathcal{E}(x_1^2) + \frac{4}{n+2}\frac{n}{n+1}\,\mathcal{E}(x_1 x_2) + \frac{4}{n+2}\frac{1}{n+1}\,\mathcal{E}(x_1 x_3),$$

and in general

$$\mathcal{E}(m_1 m_{1+d}) = \frac{n-d}{n+d}\,\mathcal{E}(x_1^2) + \frac{2d!}{(n+d)!}\sum_{t=1}^{d}\left[\frac{(n-d-t-1)!}{(d-t)!}\,\mathcal{E}(x_1 x_{1+t})\right].$$

The required correlation is now given by

$$\rho(m_1, m_{1+d}) = [\mathcal{E}(m_1 m_{1+d}) - \mathcal{E}^2(m_1)]/\mathrm{var}\,(m_1).$$

In the case of a normal parent population the first two moments and product-moments of order statistics have been tabulated for $n \leqslant 10$ (Godwin, 1949), so that ρ can be calculated for $n \leqslant 5$ (see Table 1). Explicit expressions for ρ can also be given for $n \leqslant 3$ from the exact results for order statistics up to $n = 6$ (Jones, 1948; Godwin, 1949).

3. Moving ranges

To obtain the correlation between two ranges in overlapping samples we note that

$$\mathcal{E}(w_1 w_{1+d}) = \mathcal{E}(m_1 - m_1')(m_{1+d} - m_{1+d}').$$

For a symmetrical population this reduces to

$$\mathcal{E}(w_1 w_{1+d}) = 2[\mathcal{E}(m_1 m_{1+d}) - \mathcal{E}(m_1 m_{1+d}')].$$

$\mathcal{E}(m_1 m_{1+d}')$ can again be expressed in terms of the expectations of products of order statistics in samples of $(n+d)$. It will be convenient to write x_j' in place of $x_{n+d+1-j}$, to denote the jth smallest value of x. Then

$$\mathcal{E}(m_1 m_{1+d}') = \sum_{i,j=1}^{1+d} p_{ij}\,\mathcal{E}(x_i x_j'),$$

where the coefficients p_{ij} may be obtained as follows.

Split up the $(n+d)$ values, arranged in order of time, into three groups [a], [b] and [c], comprising respectively the first d, the central $(n-d)$ and the last d of the $x_{(i)}$. Then p_{ij} is the joint probability of

$$\max\,[a+b] = x_i \quad \text{and} \quad \min\,[b+c] = x_j',$$

where, for example, $\max\,[a+b]$ stands for the largest $x_{(i)}$ in the first $(a+b)$. Now the joint probability of the $(i-1)$ largest $x_{(i)}$ lying in [c] and the $(j-1)$ smallest in [a] is

$$\frac{(d!)^2\,(n+d-i-j+2)!}{(d-i+1)!\,(d-j+1)!\,(n+d)!}.$$

We next require x_i to lie in $[a+b]$ and x_j' in $[b+c]$. According as x_i falls into [a] or [b] this joint probability is respectively

$$\frac{(d-j+1)\,(n-i+1)}{(n+d-i-j+2)\,(n+d-i-j+1)} \quad \text{or} \quad \frac{(n-d)\,(n-i)}{(n+d-i-j+2)\,(n+d-i-j+1)}.$$

Hence we have

$$p_{ij} = \frac{(d!)^2\,(n+d-i-j)!}{(d-i+1)!\,(d-j+1)!\,(n+d)!}\,[(d-j+1)\,(n-i+1)+(n-d)\,(n-i)].$$

The required correlation coefficients are now readily obtained and, for a normal parent, are given in Table 1 up to $n = 5$.

Table 1. *Correlations between maxima and ranges in overlapping*
samples of n having (n − d) values in common

n	d	Maxima	Ranges
2	1	0·4264	0·2239
3	1	0·5990	0·5395
	2	0·2743	0·2062
4	1	0·6923	0·6651
	2	0·4309	0·3942
	3	0·2028	0·1732
5	1	0·7506	0·7353
	2	0·5322	0·5097
	3	0·3374	0·3149
	4	0·1611	0·1456

4. THE EFFICIENCY OF MEAN MOVING RANGES

When moving ranges are employed in quality control it is convenient to use as the preliminary estimate of σ, calculated from N observations, the statistic

$$\omega_1 = \sum_{i=1}^{N-n+1} w_i/(N-n+1).$$

More generally, a mean moving range ω_r may be constructed based on every rth sample.
The efficiency E of ω_r defined by

$$E = \text{var } s'/\text{var } \omega_r,$$

where

$$s' = \Gamma\left(\frac{N-1}{2}\right)\sqrt{\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]}\Big/\left[\sqrt{2}\,\Gamma\left(\frac{N}{2}\right)\right],$$

is given in Table 2 in a number of cases with $n \leqslant 5$. For purposes of comparison we list E also for the ordinary mean range estimator $\bar{w}$ in samples of five and the most efficient (non-overlapping) weighted mean range estimator w^* (see Grubbs & Weaver, 1947).

Table 2. *Efficiency of various mean range estimators based on samples of n from N observations*

N	ω_1				ω_2	$\dot{w}$	w^*
	$n=2$	$n=3$	$n=4$	$n=5$	$n=4$	$n=5$	—
10	0·643	0·735	0·753	0·747	0·785	0·826	0·850
20	0·623	0·730	0·767	0·772	0·772	0·773	0·799
30	0·617	0·729	0·775	0·788	0·769	0·756	0·786
40	0·614	0·729	0·778	0·796	0·768	0·748	0·778
50	0·612	0·729	0·780	0·801	0·768	0·743	0·773
60	0·611	0·729	0·782	0·804	0·767	0·740	0·769
80	0·609	0·729	0·784	0·808	0·767	0·736	0·767
100	0·608	0·729	0·785	0·811	0·766	0·734	0·762
200	0·607	0·729	0·788	0·816	0·766	0·729	0·759
∞	0·605	0·728	0·790	0·821	0·765	0·725	0·754

It will be noted that E is lowest for ω, with $n = 2$. This is, of course, the case of the mean successive difference which is, however, less influenced by trends than the statistics based on larger r.

Compared with w and w^* these statistics are seen to gain in efficiency as N increases. Their relatively low efficiency for small N is presumably due to the uneven weighting of the observations which is inherent in their construction.

For fixed n all the statistics ω, w and w^* tend to normality with increasing N. This is obvious for w and w^* and has, in fact, been proved explicitly for ω, (Hoel, 1946).

5. RUNS OF EQUAL RANGES

Two ranges in samples of n, with $(n-d)$ observations in common, will be equal if two of these $(n-d)$ are the extremes in the $(n+d)$ values involved. Thus *for any continuous population* the probability of equality is

$$P_{d+1} = [(n-d)(n-d-1)]/[(n+d)(n+d-1)].$$

Clearly, P_{d+1} may be interpreted as the probability of a run of $(d+1)$ or more equal ranges, so that the distribution of runs $p(r)$ is determined. It is easily shown that $p(r)$ is J-shaped, flattening out with increasing n.

The expected number of runs is given by

$$\mathscr{E}(r) = \sum_{d=0}^{n-2} P_{d+1} = (3n-2) - 2(2n-1)[\psi(2n-1) - \psi(n)],$$

where $\psi(n) = d \log \Gamma(n)\, dn$. $\mathscr{E}(r)$ is evaluated in Table 3 for $n \leqslant 20$ together with the expected number of runs of equal maxima, viz.

$$\sum_{d=0}^{n-1} (n-d)/(n+d) = n[2\psi(2n) - 2\psi(n) - 1].$$

Table 3. *Expected number of runs of equal maxima and ranges*

n	Maxima	Ranges	n	Maxima	Ranges
2	1·333	1·000	8	3·606	2·239
3	1·700	1·167	9	3·991	2·462
4	2·076	1·367	10	4·375	2·687
5	2·456	1·579	12	5·146	3·137
6	2·839	1·796	15	6·303	3·815
7	3·222	2·017	20	8·232	4·947

REFERENCES

DUNCAN, A. J. (1952). *Quality Control and Industrial Statistics*. Chicago: Richard D. Irwin.
GODWIN, H. J. (1949). *Ann. Math. Statist.* **20**, 279.
GRANT, E. L. (1946). *Statistical Quality Control*. New York and London: McGraw-Hill.
GRUBBS, F. E. & WEAVER, L. C. (1947). *J. Amer. Statist. Ass.* **42**, 224.
HOEL, P. G. (1946). *Ann. Math. Statist.* **17**, 475.
HOWELL, J. M. (1949). *Ann. Math. Statist.* **20**, 305.
JONES, H. L. (1948). *Ann. Math. Statist.* **19**, 270.

Censored samples from truncated normal distributions*

By A. CLIFFORD COHEN, Jr.

The University of Georgia

1. INTRODUCTION AND SUMMARY

In life-testing and response-time studies, selection procedures sometimes operate to effect a truncation at the lower end of the time scale prior to starting a test which is subsequently terminated before all sample specimens exhibit the reaction being studied. Resulting samples may thus be regarded as censored from a truncated population. The present paper is limited to censored samples from truncated normal distributions. It is related to previous studies of truncated and censored samples by Hald (1949), the author (1950, 1954), Gupta (1952), and various other writers. It is also related to a paper on life testing by Epstein & Sobel (1953), in which some of the advantages of employing censored samples to conserve time and test specimens are discussed with regard to a one-parameter exponential distribution. For samples of the types considered here, maximum-likelihood estimators (estimates) of the population mean and standard deviation are derived, and their asymptotic variances are obtained. An illustrative example is given to demonstrate the practical application of these results.

2. MAXIMUM-LIKELIHOOD ESTIMATION

The probability density (frequency) function of a normal distribution with mean, m, and standard deviation, σ, truncated on the left at a fixed terminus, x_0, may be written as

$$f(x) = (I_1 \sigma \sqrt{(2\pi)})^{-1} \exp - [(x-m)^2/(2\sigma^2)] \quad (x_0 \leqslant x \leqslant \infty), \Big\}$$
$$f(x) = 0 \qquad\qquad\qquad\qquad (x < x_0), \quad\Big\} \tag{1}$$

where $I_1 = I(\xi_1)$, is the proportion of the complete distribution retained after truncation and ξ_1 is the left terminus (truncation point) in standard units of the complete distribution, i.e.

$$\xi_1 = (x_0 - m)/\sigma, \quad I(\xi) = \int_\xi^\infty \phi(t)\,dt, \quad \phi(t) = (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}t^2). \tag{2}$$

Let N sample specimens be selected from a population distributed according to (1) and let their life spans or reaction times $\{x_i\}$ $(i = 1, 2, 3, ..., N)$ be observed and recorded until a fixed number $(n < N)$ have reacted. Let observation be discontinued upon determining x_n. Thus x_n is the greatest of the measured observations, and it is known that each of the $N - n$ censored observations exceeds that value. When x is distributed according to (1), the likelihood function for a sample of the type described is

$$P = \left\{ K(\sigma I_1 \sqrt{(2\pi)})^{-n} \exp\left[-\sum_1^n (x_i - m)^2/(2\sigma^2) \right] \right\} (I_2/I_1)^{N-n}, \tag{3}$$

where K is a constant, $\qquad \xi_2 = (x_n - m)/\sigma \quad$ and $\quad I_2 = I(\xi_2)$. $\qquad\qquad$ (4)

According to Gupta (1952), a sample in which the number of censored (unmeasured) observations is fixed, is designated as Type II censored as distinguished from Type I censored samples in which the terminal is fixed. If rather than stopping a test after n (fixed) observations have been made, it is terminated at the expiration of a fixed time, x_u, then n is a random variable and the resulting sample is of Type I. The likelihood function of a Type I censored sample differs from (3) only in that K is replaced by a different constant, and in (4), x_n is replaced by x_u. Consequently, the same estimators are applicable in both cases except for the trivial case of a Type I censored sample in which $n = 0$. In this latter situation, the estimators would not be defined.

Taking logarithms of (3), differentiating and equating to zero, we have

$$\frac{\partial L}{\partial m} = -\frac{N\phi_1}{\sigma I_1} + \frac{1}{\sigma}\sum_1^n \left(\frac{x_i - m}{\sigma}\right) + \frac{N-n}{\sigma}\frac{\phi_2}{I_2} = 0, \Bigg\}$$
$$\frac{\partial L}{\partial \sigma} = -\frac{N\xi_1\phi_1}{\sigma I_1} - \frac{n}{\sigma} + \frac{1}{\sigma}\sum_1^n \left(\frac{x_i - m}{\sigma}\right)^2 + \frac{N-n}{\sigma}\frac{\xi_2\phi_2}{I_2} = 0, \Bigg\} \tag{5}$$

where $L = \log P$ and $\phi_i = \phi(\xi_i)$ $(i = 1, 2)$.

Let Z_i designate the reciprocal of Mill's ratio, i.e.

$$Z_i = \phi_i/I_i = e^{-\frac{1}{2}\xi_i^2}\Big/\int_{\xi_i}^{\infty} e^{-\frac{1}{2}t^2}dt \quad (i = 1, 2). \tag{6}$$

Substitute (6) into (5) and simultaneously substitute $m = x_i - \sigma\xi_i$, which follows from (2). After simplifying, we obtain the estimating equations

$$\nu_1\,w = \{(N/n)Z_1 - [(N-n)/n]Z_2 - \xi_1\}(\xi_2 - \xi_1), \tag{a}$$
$$s^2\,w^2 = \{1 + (N/n)\xi_1 Z_1 - [(N-n)/n]\xi_2 Z_2 - [(N/n)Z_1 - ((N-n)/n)Z_2]^2\}(\xi_2 - \xi_1)^2. \tag{b} \tag{7}$$

where w is the restricted sample range, ν_k is the kth sample moment about the left terminus and s^2 is the sample variance, i.e.

$$w = x_n - x_0, \quad \nu_k = \sum_1^n (x_i - x_0)^k/n, \quad s^2 = \nu_2 - \nu_1^2. \tag{8}$$

The two equations of (7) can be solved simultaneously for $\hat{\xi}_1$ and $\hat{\xi}_2$ as illustrated in §4. With these estimates determined, estimates of the mean and standard deviation follow from (2) and (4) as

$$\hat{\sigma} = w/(\hat{\xi}_2 - \hat{\xi}_1) \quad \text{and} \quad \hat{m} = x_0 - \hat{\sigma}\hat{\xi}_1. \tag{9}$$

The standard maximum-likelihood symbol ($\wedge$) serves to distinguish estimates from parameters estimated.

Estimating equations (7) can be expressed in a form that is analogous to corresponding equations previously given by the author (1950) for doubly truncated normal samples. However, Z_i as used here is not defined quite the same as in the earlier paper. Here $Z_i = Z(\xi_i)$ is the reciprocal of Mill's ratio and is a function of ξ_i only, whereas for the doubly truncated cases, Z_1 and Z_2 were each defined as functions of both ξ_1 and ξ_2.

When truncation is on the right and censoring is on the left, it is merely necessary to reverse the signs of all observations and proceed as for the case discussed above since with $f(x)$ truncated on the right, $f(-x)$ will be truncated on the left.

3. VARIANCES OF ESTIMATES

The variance-covariance matrix of $(\hat{m}, \hat{\sigma})$ is derived from the second-order partial derivatives of L. We let $\phi_{11}(\xi_1, \xi_2)$, $\phi_{12}(\xi_1, \xi_2)$ and $\phi_{22}(\xi_1, \xi_2)$ designate stochastic limits of $-\dfrac{\sigma^2\partial^2 L}{n\,\partial m^2}$, $-\dfrac{\sigma^2}{n}\dfrac{\partial^2 L}{\partial m\,\partial\sigma}$ and $-\dfrac{\sigma^2\partial^2 L}{n\,\partial\sigma^2}$ as $n \to \infty$. Thus we have

$$\phi_{11}(\xi_1, \xi_2) = 1 - (N/n)Z_1(Z_1 - \xi_1) + [(N-n)/n]Z_2(Z_2 - \xi_2),$$
$$\phi_{12}(\xi_1, \xi_2) = (N/n)Z_1[1 - \xi_1(Z_1 - \xi_1)] - [(N-n)/n]Z_2[1 - \xi_2(Z_2 - \xi_2)], \tag{10}$$
$$\phi_{22}(\xi_1, \xi_2) = 2 + (N/n)\xi_1 Z_1[1 - \xi_1(Z_1 - \xi_1)] - [(N-n)/n]\xi_2 Z_2[1 - \xi_2(Z_2 - \xi_2)].$$

The asymptotic variances and the covariance are then given as

$$\text{var}(\hat{m}) \sim [\hat{\sigma}^2/n][\hat{\phi}_{22}/(\hat{\phi}_{11}\hat{\phi}_{22} - \hat{\phi}_{12}^2)],$$
$$\text{var}(\hat{\sigma}) \sim [\hat{\sigma}^2/n][\hat{\phi}_{11}/(\hat{\phi}_{11}\hat{\phi}_{22} - \hat{\phi}_{12}^2)], \tag{11}$$
$$\text{cov}(\hat{m}, \hat{\sigma}) \sim [\hat{\sigma}^2/n][-\hat{\phi}_{12}/(\hat{\phi}_{11}\hat{\phi}_{22} - \hat{\phi}_{12}^2)],$$

where $\hat{\phi}_{ij}$ is written for $\phi_{ij}(\hat{\xi}_1, \hat{\xi}_2)$.

It subsequently follows that the correlation between $\hat{m}$ and $\hat{\sigma}$ is given by

$$\rho_{\hat{m},\hat{\sigma}} \sim -\hat{\phi}_{12}/\sqrt{(\hat{\phi}_{11}\hat{\phi}_{22})}. \tag{12}$$

As given above, (10), (11) and (12) are for the Type II censored case in which N and n are fixed while x_n is a random variable. These same results also apply in the Type I case in which N and x_u are fixed while n is a random variable, if we replace n, N/n and $(N-n)/n$ by their expected values,

$$E(n) = N(I_1 - I_2)/I_1, \quad E(N/n) = I_1/(I_1 - I_2) \quad \text{and} \quad E[(N-n)/n] = I_2/(I_1 - I_2).$$

4. AN ILLUSTRATIVE EXAMPLE

One hundred units of a certain type of electronic device are selected for life testing from a group which has survived 600 hr. of prior service. The number of units which expired prior to the time of selection is unknown. The total life in hours is recorded as each of the first ninety units of the sample expires, and

the test is terminated as soon as the ninetieth unit expires. For purposes of this illustration, we consider logarithms $\{x\}$ of the life spans to be normally distributed so that for the sample selected, we have:

$$N = 100, \quad n = 90, \quad x_0 = \log 600 = 2 \cdot 778151, \quad w = 0 \cdot 235276, \quad \nu_1 = 0 \cdot 12560147,$$
$$\bar{s}^2 = 0 \cdot 00348667813, \quad \nu_1 w = 0 \cdot 533847354 \quad \text{and} \quad \bar{s}^2 w^2 = 0 \cdot 062987824.$$

To solve estimating equations (7), we employ an iterative procedure with initial approximations read from a chart recently given by the writer (1954) for graphically solving estimating equations of doubly truncated samples. By this procedure, information provided by knowledge of the number of censored observations is neglected until subsequent iterations. Thereby we obtain $\xi_1^{(0)} = -1 \cdot 60$ and $\xi_2^{(0)} = 1 \cdot 21$. To improve these initial approximations, we substitute $\xi_1^{(0)} = -1 \cdot 600$ into (7b) and by inverse interpolation obtain $\xi_2^{(1)} = 1 \cdot 302$. This value is then substituted into (7a) and we find $\xi_1^{(1)} = -1 \cdot 663$. Repeating the cycle, we find $\xi_2^{(2)} = 1 \cdot 287$ and $\xi_1^{(2)} = -1 \cdot 626$. In the notation employed here $\xi_i^{(j)}$ is the jth approximation to ξ_i. Closer approximations to the required estimates could be reached by continuing the iterations described above through additional cycles. However, greater improvement can be achieved with less labour if we take advantage of the fact that the iterants already obtained locate two points, $P_1(\xi_1^{(0)}, \xi_2^{(1)})$ and $P_3(\xi_1^{(1)}, \xi_2^{(2)})$, which lie on the curve defined by (7b), and two points, $P_2(\xi_1^{(1)}, \xi_2^{(1)})$ and $P_4(\xi_1^{(2)}, \xi_2^{(2)})$, which lie on the curve defined by (7a). We approximate the two curves in this vicinity by straight lines through these two pairs of points, and coordinates of their intersection provide improved approximations to the required estimates.

Since P_1 and P_2 each have the same ordinate, and since P_3 and P_4 likewise have the same ordinate, the required coordinates are readily determined by interpolation as summarized below.

ξ_2	ξ_1 by (7a)	ξ_1 by (7b)	Diff.
1·302	−1·663	−1·600	−0·063
1·293	−1·640	−1·640	0
1·287	−1·626	−1·663	+0·037

Thus we have $\hat{\xi}_1 = -1 \cdot 640$ and $\hat{\xi}_2 = 1 \cdot 293$, and from (9) it follows that

$$\hat{\sigma}_x = 0 \cdot 080217, \quad \hat{m}_x = 2 \cdot 90971.$$

Since $x = \log T$, where T is the life in hours, we estimate the mean life in hours as $\hat{m}_T = 812 \cdot 3 \, \text{hr}$.

If additional decimal places had been required in the above results, we could have continued the iterative process described above through one or more further cycles.

Variances and covariances of the above estimates may be computed using (10) and (11). Although ordinary tables of normal curve areas and ordinates are adequate for this purpose, tables provided by Sampford (1952) greatly facilitate the computations. Sampford tabulated

$$\lambda = Z(Z - \xi) \quad \text{and} \quad \zeta = Z[1 - \xi(Z - \xi)],$$

with η designating the argument rather than ξ as here. In using his tables, however, a word of caution is necessary since an unfortunate printing error resulted in negative signs before some of the entries of ζ whereas all of these entries should be positive.

With $\hat{\xi}_1 = -1 \cdot 640$ and $\hat{\xi}_2 = 1 \cdot 293$, we interpolate from Sampford's tables, and using (10) compute $\hat{\phi}_{11} = 0 \cdot 87961$, $\hat{\phi}_{12} = 0 \cdot 39418$ and $\hat{\phi}_{22} = 1 \cdot 12908$. On substituting these values in (11) and (12), we compute

$$\text{var}(\hat{m}) \sim 0 \cdot 0000964, \quad \text{var}(\hat{\sigma}) \sim 0 \cdot 0000751 \quad \text{and} \quad \rho_{\hat{m},\hat{\sigma}} \sim -0 \cdot 3955.$$

REFERENCES

Cohen, A. C., Jr. (1950). Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. *Ann. Math. Statist.* **21**, 557.

Cohen, A. C., Jr. (1954). Truncated and censored samples. Tech. Rep. no. 11, Ordnance research contract DA-01-009-ORD-288. *Univ. Ga. Dep. Math.*

Epstein, B. & Sobel, M. (1953). Life testing. *J. Amer. Statist. Ass.* **48**, 486.

GUPTA, A. K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika*, **39**, 260.

HALD, A. (1949). Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Skand. AktuarTidskr*. **32**, 119.

SAMPFORD, M. R. (1952). The estimation of response-time distributions. II. *Biometrics*, **8**, 307.

The rapid calculation of χ² as a test of homogeneity from a 2 · n table

By J. B. S. HALDANE

Department of Biometry, University College, London

In genetical work we very frequently have to test a $(2 \times n)$-fold table for homogeneity. We count, for example, non-recombinants and recombinants in an experiment on linkage. The frequency of the latter is unknown or only roughly known before the experiment. Only if the values found in several families are homogeneous have we the right to assume that the value derived from the total has the precision given by its standard sampling error. Suppose our table to be

a_1	a_2	—	a_r	=	a_n		A
b_1	b_2	—	b_r		b_n		B
s_1	s_2	—	s_r	—	s_n		N

That is to say the rth sample of s_r members consists of a_r members of one class and b_r of the other, while $\Sigma a_r = A$, $\Sigma b_r = B$, $A + B = N$. The exact contribution of the rth sample to χ^2 as a test of homogeneity is $\dfrac{(a_r B - b_r A)^2}{A B s_r}$, or one of many equivalent expressions.

Suppose $A > B$. If k is the nearest integer to AB^{-1}, $(A - kB) N$ may be small, say < 0.1. If not we can take a convergent $k \ h$ to A/B, where h and k are small integers, and $(hA - kB) N$ will be small.

Then the following is an exact expression for the 'homogeneity χ^2' with $n - 1$ degrees of freedom, where h may be unity:

$$\chi_h^2 = \frac{N^2}{(h+k)^2 A B} \left[\sum_{r=1}^{n} (ha_r - kb_r)^2 s_r^{-1} - (hA - kB)^2 N^{-1} \right], \tag{1}$$

whilst the following is in error by a factor of less than $\left[1 - \dfrac{(hA - kB)}{N} \right]$:

$$\chi_h^2 = \frac{1}{hk} \left[\sum_{r=1}^{n} (ha_r - kb_r)^2 s_r^{-1} - (hA - kB)^2 N^{-1} \right]. \tag{2}$$

If A and B are of the order of 1000, $(Ba_r - Ab_r)^2$ will be an integer with about five more figures than $(ha_r - kb_r)^2$, and the saving of labour is very considerable, while for smaller values most of the calculations can be reduced to mental arithmetic. The proof is as follows.

Let
$$\chi_r^2 = \frac{(Ba_r - Ab_r)^2}{A B s_r}, \quad \text{so} \quad \chi_h^2 = \sum_{r=1}^{n} \chi_r^2.$$

Let
$$x_r = a_r - A s_r N^{-1} = B s_r N^{-1} - b_r.$$

Then
$$\Sigma x_r = 0 \quad \text{and} \quad \chi_r^2 = N^2 A^{-1} B^{-1} x_r^2 s_r^{-1}.$$

Also
$$ha_r - kb_r = (hA - kB) s_r N^{-1} + (h+k) x_r.$$

So
$$(ha_r - kb_r)^2 s_r^{-1} = (hA - kB)^2 N^{-2} s_r + 2(h+k)(hA - kB) x_r + (h+k)^2 x_r^2$$

$$= (hA - kB)^2 N^{-2} s_r + 2(h+k)(hA - kB) x_r + (h+k)^2 A B N^{-2} \chi_r^2.$$

Hence
$$\sum_{r=1}^{n} [(ha_r - kb_r)^2 s_r^{-1}] = (hA - kB)^2 N^{-1} + (h+k)^2 A B N^{-2} \chi_h^2,$$

34

whence (1) follows. But if $hA - kB = D$, then

$$A = \frac{kN+D}{h+k}, \quad B = \frac{hN-D}{h+k}, \quad \frac{N^2}{(h+k)^2 AB} = (hk)^{-1}\left[1 + \frac{(h-k)D}{hkN} - \frac{D^2}{hkN^2}\right]^{-1}.$$

whence (2) follows.

As an example I take Table 1, where a_r and b_r are the numbers of normal and *vestigial Drosophila melanogaster* counted in eleven bottles in a class experiment where a $3:1$ ratio was expected on Mendelian theory, but *vestigial* was known to be somewhat inviable. The question arose whether the mortality of recessives had been significantly different in different bottles. Clearly $h = 1, k = 6$, gives a good fit.

Table 1

a_r	b_r	s_r	$a_r - 6b_r$	$(a_r - 6b_r)^2/s_r$
25	1	26	$+19$	13·885
80	15	95	-10	1·053
38	12	50	-34	23·120
52	8	60	$+4$	0·267
9	0	9	$+9$	9·000
21	7	28	-21	15·750
33	6	39	-3	0·231
24	2	26	$+12$	5·538
30	7	37	-12	3·892
51	7	58	$+9$	1·397
56	3	59	$+38$	24·475
419	68	487	$+11$	98·608

Here $h = 1, k = 6, A = 419, B = 68, N = 487$. $\sum_{r=1}^{n} (a_r - 6b_r)^2 s_r^{-1} - (A - 6B)^2 N^{-1} = 98·360$. So formula (2) gives $\chi^2 = 16·393$, formula (1) $\chi^2 = 16·709$. There are 10 degrees of freedom, so $0·09 > P > 0·08$, and the data are not significantly heterogeneous. Had the expression $\sum[(Ba_r - Ab_r)^2 s_r^{-1}](AB)$ been used, the first summand would have been 1281^2 26. In practice only two decimal places were used in the sum, and formula (2) was used, the whole calculation being completed on the blackboard in about ten minutes.

It is possible to justify formula (2) verbally as follows. $\sum[(h_r a_r - kb_r)^2 n_r^{-1}]$ is the value of χ^2 for n degrees of freedom, if we assume $\mathscr{E}(a_r) = ks_r (h+k)$; while $(hA - kB)^2 N^{-1}$ is the value of χ^2 for A and B on the same assumption. Subtracting this we obtain χ_h^2 with $n-1$ degrees of freedom. In fact, the dissection of a χ^2 into components is only valid when all samples are large, and in this case leads to an error of 2%. While this dissection is often justifiable, it is desirable to know the magnitude of its error.

The 'Inefficiency' of the sample median for many familiar symmetric distributions*

By JOHN T. CHU, *University of North Carolina*

1. A LOWER BOUND FOR THE VARIANCE OF THE MEDIAN

If the reciprocal of the (asymptotic) variance of an estimate is taken as a measure of its (asymptotic) efficiency, the sample median $\tilde{x}$ is often (asymptotically) less efficient than the sample mean x, for many symmetric distributions familiar to statisticians. In fact, for a symmetric distribution having its absolute maximum at the point of symmetry, if $\tilde{x}$ is asymptotically less efficient than x, then quite often $\tilde{x}$ is never so efficient as x, with the possible exception of very small samples. To show these facts, we first derive a very simple, yet sharp, lower bound for the variance of the sample median.

Suppose that $F(x)$ and $f(x)$ are the cumulative distribution function and the probability density function of a certain continuous distribution and $\tilde{x}$, respectively, with respect to $\mu = \xi$ and $f(\xi) > 0$ for all x. Let $\tilde{x}$ be the median of a sample of size $2n+1$, then writing $C_n = (2n+1)!/(n!)^2$,

$$\operatorname{var}\tilde{x} = \int_{-\infty}^{\infty} (x-\xi)^2 C_n[F(x)]^n[1-F(x)]^n f(x)\,dx$$

$$= \int_0^1 (x-\xi)^2 C_n F^n(1-F)^n\,dF$$

$$\geq [f(\xi)]^{-2}\int_0^1 (F-\tfrac12)^2 C_n F^n(1-F)^n\,dF$$

$$= \{4[f(\xi)]^2(2n+3)\}^{-1}. \tag{1}$$

The equality holds for a rectangular distribution.

2. Examples

It is well known that for normal and rectangular distributions $\tilde{x}$ is more efficient than $\bar{x}$. We shall show that this is true for many other familiar symmetric distributions.

(2·1) *Triangular distribution.* $f(x) = 1 - |x|$, $|x| < 1$. The variance of x is $\tfrac16(2n+1)^{-1}$ and from (1) it follows that $\tilde{x}$ is less efficient than x for samples of sizes $2n+1$ provided $n > 1$. Direct computation shows that this is also true if $n = 1$.

(2·2) *t-Distribution.* $f(t) = A(1 + t^2/k)^{-\frac12(k+1)}$, where $A^{-1} = \Gamma(\tfrac12 k)(k\pi)^{\frac12}/\Gamma[\tfrac12(k+1)]$. Since $\sigma_x^2 = k/(k-2)$, it follows that for a t-distribution with k degrees of freedom, $\tilde{x}$ is less efficient than x if (but not necessarily only if)

$$\frac{\pi(k-2)\,\Gamma^2(\tfrac12 k)}{4\Gamma^2(\tfrac12 k+\tfrac12)} > \frac{2n+3}{2n+1}.$$

For both $k = 2m$ and $k = 2m+1$, the left-hand side of the above inequality is an increasing function of m. Computation shows, e.g. that the inequality holds if $k \geq 5$ and $n \geq 25$.

(2·3) *Symmetric β-distribution.*

$$f(x) = \frac{\Gamma(2p)}{\Gamma^2(p)} x^{p-1}(1-x)^{p-1} \quad (0<x<1;\ p>0).$$

For this distribution $\sigma_x^2 = \tfrac14(2p+1)^{-1}$. Hence $\tilde{x}$ is less efficient than $\bar{x}$ if

$$2^{4p-4}(2p+1)\,\Gamma^4(p)/\Gamma^2(2p) > \frac{2n+3}{2n+1}.$$

The left-hand side becomes smaller if p is replaced by $p+1$, and tends to $\tfrac12\pi$ as $p \to \infty$. So it has a lower bound $\tfrac12\pi$. Hence the inequality holds for every $p>0$ and $n \geq 2$.

(2·4) *Cauchy type distribution.* This is defined to be of the type $f(x) = C_\alpha(1 + x^2)^{-\alpha}$ $(-\infty < x < \infty; \alpha > 1)$. If $\alpha = 2$, we obtain the well-known Cauchy distribution for which $\tilde{x}$ is infinitely more efficient than $\bar{x}$. It would be interesting to examine whether or not $\bar{x}$ becomes more efficient as α increases. Now x has a finite variance only if $\alpha > \tfrac32$. C_α and $\operatorname{var} x$ can be obtained by using contour integration (see Whittaker & Watson, 1952, p. 118). It follows that $\tilde{x}$ is less efficient than x if

$$\frac{y^2 \sin 3y}{\sin^5 y} > \frac{2n+3}{2n+1},$$

where $y = \pi/\alpha$. The left-hand side is a decreasing function of y, and so an increasing function of α. The least values of α for which the left-hand side is equal to 5, 3 and 1 (the maximum and minimum of the right-hand side), are found to be 4·65 and 3·75 approximately.

3. Remarks

(i) $\bar{x}$ is not more efficient than $\tilde{x}$ for all symmetric distributions. When the parent population has a Laplace distribution, for example, $\tilde{x}$ is more efficient for all samples of odd sizes (Chu & Hotelling).

(ii) If $f(x)$ satisfies certain continuity conditions, $\tilde{x}$ has an asymptotically normal distribution and the asymptotic variance is $\{4[f(\xi)]^2(2n+1)\}^{-1}$. If the sample size, therefore, is not too small, the asymptotic variance is for all practical purposes a lower bound for the variance of $\tilde{x}$. And if $\tilde{x}$ is asymptotically less efficient than $\bar{x}$, then $\tilde{x}$ is less efficient than $\bar{x}$ for all samples whose sizes are not too small.

REFERENCES

Chu, J. T. & Hotelling, H. The moments of the sample median (to be published).

Whittaker, E. T. & Watson, G. N. (1952). *Modern Analysis*. Cambridge University Press.

A simple method of calculating the exact probability in 2 × 2 contingency tables with small marginal totals

By P. H. LESLIE

Bureau of Animal Population, Department of Zoological Field Studies, Oxford

It frequently happens that in the case of data which may be arranged in the form of a 2 × 2 contingency table, we wish to determine the exact probability of obtaining the observed result. In the method which appears to be adopted usually, we have to calculate the value of a number of expressions involving factorials. Although the actual calculations are not difficult, they become somewhat tedious even with the help of a table of log factorials, and a simple method of obtaining the exact solution may therefore be of practical use. The following method is convenient when the marginal totals are small (and usually it is only in such cases that we need to calculate the exact solution), and only a table of the binomial coefficients up to, say, $n = 20$ is required. Tables of these coefficients seem to be published up to $n = 12$, e.g. Barlow, and Milne-Thomson & Comrie, four-figure tables; but they may be extended once and for all to higher values very simply by means of Pascal's triangle.

Suppose we write any 2 × 2 table in the standard form:

x		n_A
		$N - n_A$
n_B	$N - n_B$	N

where N is the total number of observations, and the marginal totals fulfil the following inequalities,

$$n_A \leqslant N - n_A, \quad n_B \leqslant N - n_B \quad \text{and} \quad n_B \leqslant n_A;$$

so that n_B is the smallest marginal total, and x the number of observations in the cell with the smallest expectation. Then the set of possible results which is compatible with the given marginal totals is obtained by giving x in turn the successive values $0, 1, 2, ..., n_B$, and the probability of obtaining a particular value of x is given by the appropriate term of the hypergeometric distribution,

$$P(x) = \binom{n_A}{x} \binom{N - n_A}{n_B - x} \bigg/ \binom{N}{n_B}.$$

If, as a first step, we write down opposite the values of $x = 0, 1, 2, ..., n_B$, starting at $x = 0$, the successive binomial coefficients for $n = n_A$, which we will call column a_x; and if, secondly, starting at the bottom opposite $x = n_B$, we write down in reverse order the binomial coefficients for $n = N - n_A$, and call this column b_x; then the product column $a_x b_x$ gives the successive terms of the distribution we require. The total of this last column,

$$\sum_x a_x b_x = \frac{N!}{n_B! (N - n_B)!},$$

which serves as a check if it is needed. The calculation of the exact probability $\sum_0^x P(x)$ or $\sum_x^{n_B} P(x)$ (which-ever tail of the distribution is appropriate in the particular case), then follows very quickly on an ordinary calculating machine; in fact, the individual product terms, $a_x b_x$, need not necessarily be written down.

Merely as an arithmetical example, suppose we had the following table, where $n_A = 9$, $n_B = 8$, $N = 20$ and $x = 7$,

7	2	9
1	10	11
8	12	20

Then the two essential columns are the binomial coefficients for $n = 9$, forming column a_x, and those for $n = 11$, forming column b_x. Thus

x	a_x	b_x
0	1	165
1	9	330
2	36	462
3	84	462
4	126	330
5	126	165
6	84	55
7	36	11
8	9	1

Then, the sum of the product column, $\Sigma a_x b_x = 125{,}970$, which can be obtained directly on the machine without writing down the individual terms. Since the observed number, $x = 7$, in the cell with the smallest expectation is greater than expectation, we would require in this case the right-hand tail of the distribution. Thus

$$(36 \times 11) + (9 \times 1) = 405,$$

and

$$\sum_{7}^{8} P(x) = 405/125970 = 0.003215.$$

A test for a change in a parameter occurring at an unknown point

By E. S. PAGE

Department of Mathematics, University of Durham

1. Introduction and summary

Consider a sample of independent observations in the order in which they were obtained, $x_1 \ldots x_n$; it is sometimes required to test the null hypothesis that all the observations are drawn from the same population with distribution function $F(x \mid \theta)$ against the alternative that $x_1, \ldots, x_m$ come from $F(x \mid \theta)$, and $x_{m+1}, \ldots, x_n$ from $F(x \mid \theta')$ $(\theta' \neq \theta)$. If m is known this is a straightforward problem of comparing two samples. In this paper we suppose that m is unknown; this raises new problems. A test is proposed for a case where θ is known and some comments are made on the problems presented by other cases.

2. One-sided case: θ known

Suppose that the initial value, θ, is known. One possibility for a test is to regard all the observations as a single sample and to use the best test that all the observations are from $F(x \mid \theta)$ against the alternative that all are from $F(x \mid \theta')$ for some $\theta' \neq \theta$. Such a test cannot be expected to be very powerful if the change occurs late in the sample; the few observations on the new parameter value would be obscured by the many on the old parameter. Since the problem of detecting a change in a parameter is important in controlling the quality of the output from a continuous production process, it is reasonable to investigate whether the methods of process inspection schemes can provide useful tests for the case in which we are interested.

A process inspection scheme for detecting a change in one direction in the parameter was given by the author in an earlier paper (Page, 1954). We suppose throughout that the parameter under consideration is the mean of the distribution unless the contrary is stated, and in this section we further assume that the value, θ, at the start of the observations is known. The scheme consisted of recording the cumulative sums $S_r = \sum_{i=1}^{r} (x_i - \theta)$, $S_0 = 0$, and taking action to rectify a possible change in the parameter when $S_r - \min_{0 \leq i < r} S_i \geq h$, i.e. when the sample path rises a height h above its previous minimum value; this

procedure can be displayed clearly on a chart. If there is no change in θ, the mean path of the cumulative sum is horizontal, while if an increase in θ occurs the new mean path has positive slope so that the above criterion would be satisfied without too much delay. The significance test suggested by this procedure is as follows:

I. *Given the observations* $x_1 \ldots x_n$. *It is required to test the hypothesis that the mean is constantly* θ.

Use as the test statistic $m = \max_{0 \leqslant r \leqslant n} \{S_r - \min_{0 \leqslant i < r} S_i\}$, *where* $S_r = \sum_{j=1}^{r} (x_j - \theta)$, $S_0 = 0$, *taking large values as significant, i.e. reject the hypothesis if* $m \geqslant h$.

It was shown in the paper cited that the properties of the corresponding process inspection scheme depended upon the characteristics of linear sequential tests; as the test I is a truncated form of the process inspection scheme it is to be expected that in general the properties of the test will be difficult to evaluate. A special case that is tractable is where the observations are nought-or-one binomial variables. Accordingly, we consider a test for the general case using binomial variables.

II. *Given the observations* $x_1 \ldots x_n$. *It is required to test the hypothesis that the mean is constantly* θ. *Let* $y_i = a$ *if* $x_i - \theta \geqslant 0$ *and* $y_i = -b$ *if* $x_i - \theta < 0$, *and choose* a, b (>0) *so that* $E(y_i \mid \theta) = 0$ $(i = 1, \ldots, n)$.

Use as the test statistic $m = \max_{0 \leqslant r \leqslant n} \{S_r - \min_{0 \leqslant i < r} S_i\}$, *where* $S_r = \sum_{j=1}^{r} y_j$, $S_0 = 0$, *taking large values as significant, i.e. reject the hypothesis if* $m \geqslant h$.

For simplicity we shall consider only the case where the distribution of the x_i is symmetrical, so that we can take $a = b = 1$; hence $y_i = \operatorname{sgn}(x_i - \theta)$. In order to evaluate the properties of the test let $m_r = S_r - \min_{0 \leqslant i < r} S_i$ and let $p_{r,i}$ be the probability that $m_r = i$ $(i = 0, 1, \ldots, h-1)$ and that $m_s < h$ for all s, $1 \leqslant s < r$. Then

$$1 - \sum_{i=1}^{h-1} p_{n,i} \tag{1}$$

is the probability that the null hypothesis is rejected. Let $\operatorname{prob}(y_i = 1)$ be $p = 1 - q$. By considering the result of the next observation we have the relations

$$\left.\begin{aligned}
p_{r+1,0} &= q(p_{r,0} + p_{r,1}), \\
p_{r+1,i} &= p \cdot p_{r,i-1} + q \cdot p_{r,i+1} \quad (1 \leqslant i < h-1), \\
p_{r+1,h-1} &= p \cdot p_{r,h-2}.
\end{aligned}\right\} \tag{2}$$

In matrix notation we have

$$\mathbf{p}_{r+1} = \mathbf{P} \cdot \mathbf{p}_r, \tag{3}$$

where $\mathbf{P}$ is the square matrix,

$$\mathbf{P} = \begin{pmatrix} q & q & 0 & 0 & \ldots & 0 & 0 \\ p & 0 & q & 0 & \ldots & 0 & 0 \\ 0 & p & 0 & q & \ldots & 0 & 0 \\ \multicolumn{7}{c}{\dotfill} \\ 0 & 0 & 0 & 0 & \ldots & p & 0 \end{pmatrix}. \tag{4}$$

Initially $p_{0,0} = 1$, $p_{0,i} = 0$ $(i \neq 0)$. In this formulation we are implicitly using the fact that the m_r are variables in a Markov chain with h live states. Clearly

$$\mathbf{p}_r = \mathbf{P}^r \cdot \mathbf{p}_0. \tag{5}$$

The expression for $\mathbf{p}_r$ may alternatively be written in terms of the latent roots and vectors, but the simplicity of the matrix $\mathbf{P}$ makes it quite convenient to use (4) for calculation. On the null hypothesis $p = \frac{1}{2}$ constantly. Table 1 shows values of h and the sample size n for which the probabilities of errors of the first kind are at most α, where $\alpha = 0.05$ and 0.01. In order to ensure that the Type I errors are at most α for non-tabular values of n the larger value of h should be taken. For larger values of n rough interpolation will provide a sufficiently accurate value of h.

The power of the test II depends both on the value of p after the change and the position of the change. If $\operatorname{prob}(y_i = +1)$ is constantly equal to p so that the change can be considered as having occurred immediately, the probability that the null hypothesis is rejected in a sample of n observations is given by equations (1) and (5) with $r = n$. If the change occurs after the kth observation the value of $\mathbf{p}_n$ to be used in (1) is given by

$$\mathbf{p}_n = \mathbf{P}_2^{n-k} \cdot \mathbf{P}_1^k \cdot \mathbf{p}_0,$$

where $\mathbf{P}_1$, $\mathbf{P}_2$ are the matrix $\mathbf{P}$ with $p = \frac{1}{2}$, $p = p$, respectively.

Table 1. *Values of n and h*

x = 0.05				x = 0.01	
n	h	n	h	n	h
21	10	75	19	20	12
26	11	83	20	27	14
31	12	91	21	35	16
36	13	100	22	43	18
41	14	119	24	53	20
47	15	139	26	64	22
54	16	161	28	76	24
60	17	185	30	89	26
67	18			103	28
				118	30

In Table 2 the power of the test II is compared with the power of the simple binomial test with approximately the same probability of Type I errors for a sample of 50 observations. For test II to have probability of Type I errors just less than 0·05 we need $h = 16$. The corresponding single sample test is 'Reject the null hypothesis if more than 31 of the y_i are positive', i.e. if $\sum_{i=1}^{50} y_i > 12$. The loss of power from using Test II instead of the single sample test is remarkably small.

Table 2. *Powers of the tests for different p*

p	Test II	Single sample test
0·50	0·039	0·032
0·55	0·136	0·127
0·60	0·336	0·336
0·65	0·609	0·622
0·70	0·844	0·859
0·75	0·964	0·971
0·80	0·996	0·997

The same two tests are contrasted in Table 3, where their powers are shown for different positions of the change from $p = 0.5$ to $p = 0.75$. Here, however, the test II has an appreciably greater power than the single-sample test when the change occurs near the middle of the set of observations. Also shown in Table 3 are the powers of the single-sample test on the last $50 - m$ observations, when it is known that the change has occurred immediately before the mth observation. The differences between the power of this test and that of test II gives an indication of what is lost from the ignorance of the position of change.

In order to illustrate the test we give an example constructed from tables of random normal deviates. A sample of forty observations was constructed, the first twenty having mean 5 and unit variance, and the last twenty having mean 6 and unit variance; these are shown in Table 4. Suppose that it is required to test the hypothesis that the mean is constantly 5 against the alternatives that an increase in the mean has occurred within the sample. The observations, x_i, are shown in Table 4 together with $y_i = \mathrm{sgn}\,(x_i - 5)$, and the value taken by $S_r - \min S_i$.

The greatest value, h, of $S_r - \min S_i$ in the sample of 40 is 17, which approaches the 1 % significance point given in Table 1 (for $n = 40$, the approximate 5% point is $h = 14$, the approximate 1% point is $h = 18$). This significance level can be compared with that obtained from other tests applied to the

Table 3. *Powers of the tests for different positions of the change*

m	Test II	Single-sample test on whole sample	Single-sample test, m known
0	0·964	0·971	0·971
10	0·906	0·864	0·946
20	0·733	0·625	0·894
30	0·398	0·330	0·618
40	0·122	0·122	0·244
50	0·039	0·032	0·032

Table 4. *Artificial sampling experiment*

Observation no.	1	2	3	4	5	6	7	8	9	10
Value of x_i	3·95	5·96	6·22	5·58	4·02	4·97	3·46	4·29	4·65	5·66
$y_i = \mathrm{sgn}\,(x_i - 5)$	−1	+1	+1	+1	−1	−1	−1	−1	−1	+1
$S_r - \min S_i$	0	1	2	3	2	1	0	0	0	1

Observation no.	11	12	13	14	15	16	17	18	19	20
Value of x_i	5·44	5·91	4·98	3·58	5·26	3·98	4·19	6·66	6·05	5·97
$y_i = \mathrm{sgn}\,(x_i - 5)$	+1	+1	−1	−1	+1	−1	−1	+1	+1	+1
$S_r - \min S_i$	2	3	2	1	2	1	0	1	2	3

Observation no.	21	22	23	24	25	26	27	28	29	30
Value of x_i	7·14	6·22	4·76	6·60	5·72	4·88	5·44	5·03	5·66	5·56
$y_i = \mathrm{sgn}\,(x_i - 5)$	+1	+1	−1	+1	+1	−1	+1	+1	+1	+1
$S_r - \min S_i$	4	5	4	5	6	5	6	7	8	9

Observation no.	31	32	33	34	35	36	37	38	39	40
Value of x_i	6·37	6·66	5·10	5·80	6·29	5·49	4·93	6·18	8·29	6·84
$y_i = \mathrm{sgn}\,(x_i - 5)$	+1	+1	+1	+1	+1	+1	−1	+1	+1	+1
$S_r - \min S_i$	10	11	12	13	14	15	14	15	16	17

sample. The single-sample binomial test on the y's has 26 positives, 14 negatives; on the null hypothesis, the probability of this or a larger number of positives is 0·04. The change in the mean causes the estimate of variance of the x's to be inflated, and a t-test fails to give significance. The computation required by test II is so simple that it is unnecessary to record the y's, or even the $S_r - \min S_i$. An additional advantage of the test is that it gives an indication where the change took place; the position of the last zero of $S_r - \min S_i$ is an estimate (of course, biased) of the position of change. Thus in the example we would suspect that the change had occurred near observation 17.

3. General remarks

In this section we comment briefly on some other possible methods for the problem of § 2 and related problems without investigating their properties.

Another test for a change in one direction of a parameter from a known specified value may be obtained by analogy with the standard control chart process inspection scheme. The sample is divided into a

number of subsamples of equal size, and a statistic calculated from each subsample, the hypothesis of no change is rejected unless all the statistics fall within a certain range. The properties of this test are easy to evaluate and the number of subsamples and the permissible interval for the statistics can be chosen to control the errors. The test is also easy to apply and it is frequently useful in rough work. However, the temporal ordering of the observations enters only into the division into subsamples, and it is of interest to examine whether it is advantageous to employ a slightly more complicated test of the form 'Reject the null hypothesis if any k of the statistics calculated from m consecutive subsamples fall outside an interval I, or if any one falls outside a wider interval I'' (cf. Wilkinson, 1951; Tippett, 1931).

The control chart procedure can also provide a test for the two-sided case where the change from the known value can be in either direction. A test based on the mean path of a cumulative sum similar to test I is a truncated sequential test (Rao, 1950). Another case that needs to be considered is where the initial value of the parameter is unknown.

I wish to thank Dr D. R. Cox for a number of discussions on the subject of this paper, and the Director, Mathematical Laboratory, Cambridge, for permission to use the EDSAC.

REFERENCES

PAGE, E. S. (1954). *Biometrika*, **41**, 100.
RAO, C. R. (1950). *Sankyhā*, **10**, 361.
TIPPETT, L. H. C. (1931). *The Methods of Statistics*. London: Williams and Norgate.
WILKINSON, B. (1951). *Psychol. Bull.* **48**, 156.

A paradox in statistical estimation

By ALAN STUART

Division of Research Techniques, London School of Economics

1. Sundrum (1954) has recently shown to be incorrect the intuitive idea that the more efficient of two estimators of a parameter necessarily provides the more powerful test of a hypothesis concerning that parameter. This note discusses a similar paradox which arises in a problem concerned purely with estimation: given an estimator u of a parameter θ in a multiparameter distribution, one does not necessarily improve its efficiency by substituting true parameter values into u to replace estimators of them.

2. We shall consider the case where there is only one other parameter, say μ, and where we have a consistent estimator of θ

$$t = f(s, \mu), \tag{1}$$

where s is a function of the n observations only. If μ is unknown, we reduce t to a function of the observations only by substituting for μ a consistent estimator of it, say m, giving

$$u = f(s, m). \tag{2}$$

From (1) we have, to order n^{-1},

$$V(t) = \left(\frac{\partial t}{\partial s}\right)^2 V(s). \tag{3}$$

From (2) we have the corresponding result for a function of two random variables

$$V(u) = \left(\frac{\partial u}{\partial s}\right)^2 V(s) + \left(\frac{\partial u}{\partial m}\right)^2 V(m) + 2\frac{\partial u}{\partial s}\frac{\partial u}{\partial m} C(s, m). \tag{4}$$

Since all the derivatives in (3) and (4) are to be taken at the true parameter point (θ, μ), the first term on the right of (4) is equal to (3). Thus

$$V(u) - V(t) = \left(\frac{\partial u}{\partial m}\right)^2 V(m) + 2\frac{\partial u}{\partial s}\frac{\partial u}{\partial m} C(s, m). \tag{5}$$

(5) is not generally positive, although it must be so if s and m are uncorrelated. In general, their correlation must be taken into account before the effect of substituting parameters into u can be assessed. It is the correlation term in (5) which resolves the paradox.

3. An example ... a negative ... pro... ed by the estimation of the corr... ... ...,
in a bivariate ... ... A consistent estimator is provided by the sample ... ... of
efficient r, which has a large-sample variance,

$$V(r) = \frac{1}{n}(1-\rho^2)^2.$$

(6)

Although r is the maximum likelihood estimator of ρ when all five population parameters are being
simultaneously estimated, it is not an efficient estimator when the population means and variances
are known, which is the case we shall consider. One might therefore expect to ... ... by
substituting the known population means and variances for their sample correspond... ... The new
estimator thus obtained is

$$r' = \frac{\frac{1}{n}\sum_i (x_i-\mu_1)(y_i-\mu_2)}{\sigma_1\sigma_2}.$$

(7)

We find

$$E(r') = \rho.$$

(8)

$$n^2\sigma_1^2\sigma_2^2 E(r'^2) = E.\sum_i (x_i-\mu_1)^2(y_i-\mu_2)^2 + \sum_{i\neq j}(x_i-\mu_1)(y_i-\mu_2)(x_j-\mu_1)(y_j-\mu_2)$$

$$= n\mu_{22} + n(n-1)\rho^2\sigma_1^2\sigma_2^2.$$

(9)

Since $\mu_{22} = (1+2\rho^2)\sigma_1^2\sigma_2^2$, we have from (8) and (9)

$$V(r') = \frac{1}{n}(1+\rho^2),$$

(10)

and comparing (10) with (6) we see that, far from improving the accuracy of estimation, the substitution
of true parameter values for random variables in r has multiplied efficiency by a factor $\frac{(1-\rho^2)^2}{1+\rho^2}$, which
tends to zero for large ρ^2, and is 1 only when $\rho^2 = 0$.

4. This paradoxical effect can be examined through (5) by restricting our attention to the case when
both population means are zero and both variances equal to σ^2, which remains unknown. Our estimator
(7) then becomes

$$t = \frac{\frac{1}{n}\sum_i x_i y_i}{\sigma^2},$$

and with a common population variance, we should in our correlation coefficient use a pooled estimator
of σ^2, giving

$$u = \frac{\frac{1}{n}\sum_i x_i y_i}{\frac{1}{2n}\sum_i (x_i^2+y_i^2)}.$$

We shall see below that u has the same efficiency as r.

In the notation of (1) and (2) we have

$$t = \frac{s}{\mu} \quad \text{and} \quad u = \frac{s}{m},$$

where

$$s = \frac{1}{n}\sum_i x_i y_i, \quad \mu = \sigma^2 \quad \text{and} \quad m = \frac{1}{2n}\sum_i (x_i^2+y_i^2).$$

We now require

$$V(m) = \frac{\sigma^4}{n}(1+\rho^2),$$

$$C(s,m) = \frac{2\rho\sigma^4}{n}.$$

(11)

Also

$$\left(\frac{\partial u}{\partial m}\right)_{\mu,\theta} = \left(-\frac{u}{m}\right)_{\mu,\theta} = -\frac{\rho}{\sigma^2},$$

$$\left(\frac{\partial u}{\partial s}\right)_{\mu,\theta} = \left(\frac{1}{m}\right)_{\mu} = \frac{1}{\sigma^2}.$$

(12)

Substituting (11) and (12) into (5), we find

$$V(u) - V(\theta) = \frac{\rho^2 \sigma^2}{\sigma^2 \, n}(1+\rho^2) + 2\left(\frac{1}{\sigma^2}\right)\left(-\frac{\rho}{\sigma^2}\right)\frac{2\rho\sigma^2}{n}$$

$$= \frac{\rho^2}{n}(\rho^2 - 3). \tag{13}$$

... ...

... ...

$$V(u) = \{E(u)\}^2\left\{\frac{V(s)}{\{E(s)\}^2} + \frac{V(m)}{\{E(m)\}^2} - 2\frac{C(s,m)}{E(s)\,E(m)}\right\}. \tag{14}$$

Using (8), (10) and (11) in (14), we obtain

$$V(u) = \rho^2\left\{\frac{1}{n}\frac{(1+\rho^2)}{\rho^2} + \frac{1}{n}(1+\rho^2) - 2\frac{2}{n}\right\}$$

$$= \frac{1}{n}(1-\rho^2)^2, \tag{15}$$

agreeing with (6).

REFERENCE

SUNDRUM, R. M. (1954). On the relation between estimating efficiency and the power of tests. *Biometrika*, 41, 542–4.

Cumulants of a transformed variate

By G. S. JAMES

University of Leeds

1. Suppose that x is a variate whose pth cumulant, denoted by κ_{px}, is of order ν^{-p+1}, where ν is some 'large' number. One consequence of this is that x is approximately normally distributed with a variance of order ν^{-1}. Thus if y is some well-behaved function of x, say $y = f(x)$, then y will also be approximately normally distributed with mean and variance given by

$$\kappa_{1y} = \mathscr{E}y = f(\mu)[1 + O(\nu^{-1})], \tag{1}$$

$$\kappa_{2y} = \operatorname{var} y = [f'(\mu)]^2 \operatorname{var} x[1 + O(\nu^{-1})], \tag{2}$$

where μ is the mean value of x. These are of orders ν^0 and ν^{-1} respectively.

It is by no means obvious, however, that the pth cumulant of y, denoted by κ_{py}, is of order ν^{-p+1} for general values of p, and not merely for $p = 1$ and 2. The object of this note is to show that this is true, at any rate formally. (The result will be true in reality, as well as formally, under the same sort of conditions for which (1) and (2) are really true.)

Suppose that $f(x)$ can be expanded formally in a Taylor series round μ; there is no loss of generality in taking $\mu = 0$, so that this series is

$$y = c_0 + c_1 x + c_2 x^2 + \dots, \tag{3}$$

where we suppose that $c_0, c_1, c_2, \dots$ do not depend upon ν. It is quite easy to show, by formal expansion starting from (3), that the pth moment of y about its mean is of order $\nu^{-[\frac{1}{2}p]}$, when $[k]$ denotes the integral part of k. But on conversion to cumulants the higher order terms always seem to cancel, to such an extent that κ_{py} is of order ν^{-p+1}. We now give a proof of this result.

THEOREM 1. *If a variate x possesses cumulants κ_{px} of all orders, and if $\kappa_{px} = O(\nu^{-p+1})$ $(p = 1, 2, \dots)$, and the cumulants of $y = f(x)$ are calculated on the basis of a (possibly formal) Taylor expansion (3), where the c_r do not depend upon ν, then $\kappa_{py} = O(\nu^{-p+1})$ $(p = 1, 2, \dots)$.*

Without loss of generality we may assume that $c_0 = 0$. We may then write $y = \Sigma z_r$, where $z_r = c_r x^r$ $(r = 1, 2, \dots)$. Writing $\mu'_{(p)}(r_1 \dots r_p)$ for the moment $\mathscr{E} z_{r_1} \dots z_{r_p}$ (some of the r_h may be equal) and $\kappa_{(p)}(r_1 \dots r_p)$

for the corresponding cumulant, we have, by the properties of moment-generating and cumulant-generating functions,

$$\frac{\kappa_{1y}t}{1!}+\frac{\kappa_{2y}t^2}{2!}+\dots = \log\left[1+\frac{t\delta y}{1!}+\frac{t^2\delta y^2}{2!}+\dots\right]$$

$$= \log\left[1+\frac{1}{1!}\Sigma t_r\mu'_{(1)}(r)+\frac{1}{2!}\Sigma t_r t_s\mu'_{(1)}(rs)+\dots\right]$$

$$= \frac{1}{1!}\Sigma t_r\kappa_{(2)}(r)+\frac{1}{2!}\Sigma t_r t_s\kappa_{(2)}(rs)+\dots, \tag{4}$$

where $t_1 = t_2 = \dots = t$ and the summations are over $r, s, \dots = 1, 2, \dots$. Hence

$$\kappa_{py} = \Sigma\kappa_{(y)}(r_1\dots r_p), \tag{5}$$

summed for $r_1, \dots, r_p = 1, 2, \dots$.

Thus to prove the theorem it is sufficient to show that $\kappa_{(2)}(r_1\dots r_p)$ is of order ν^{-p+1} or lower. To demonstrate this, suppose rather more generally that $x_1, \dots, x_n$ is a sample of n independent values from the x distribution, and let $z_r = c_r(\Sigma x_i)^r$ $(r = 1, 2, \dots)$. This agrees with the earlier definition of z_r when $n = 1$ (and $x_1 = x$). The z_r can be regarded as statistics of the sample $x_1, \dots, x_n$, although of course they are functionally related. Each z_r is a homogeneous polynomial symmetric function of the sample values, of degree r, and the coefficient of $x_1^{a_1}\dots x_a^{a_a}$ in z_r (with $\Sigma a_j = r$) is

$$\frac{r!}{a_1!\dots a_\alpha!}B_{(z)}(a_1\dots a_\alpha), \tag{6}$$

where

$$B_{(z)}(a_1\dots a_\alpha) = c_r. \tag{7}$$

A corresponding result holds for Fisher's statistics k_r, but with

$$B_{(k)}(a_1\dots a_\alpha) = \frac{(-)^{\alpha-1}(\alpha-1)!}{n(n-1)\dots(n-\alpha+1)}. \tag{8}$$

The factors $B_{(k)}$ and $B_{(z)}$ happen not to depend on the complete detail of the partition $(a_1\dots a_\alpha)$ of the number r; but I have shown (James, 1955) that Fisher's rules for obtaining the sampling cumulants of k-statistics, described in Kendall's book (Kendall, 1947), may be adapted to *any* system of statistics $z_1, z_2, \dots$ (z_r being a homogeneous polynomial symmetric function of degree r in the observations) by merely replacing the factors $B_{(k)}$ occurring in the evaluation of the 'pattern functions' by the factors $B_{(z)}$, where $B_{(z)}$ is the coefficient of $x_1^{a_1}\dots x_a^{a_a}$ in z_r, divided by the appropriate multinomial coefficient.

The rule of Fisher which is of paramount importance for our proof is the one which states that, in finding cumulants of k- (or z-) statistics in terms of population cumulants, we are to neglect any array which splits up into two or more disjoint blocks. Now any array which does *not* fall into disjoint blocks may be built up column by column in such a way that at each stage the new column does not form an isolated block by itself. (For example, the columns of $\begin{smallmatrix} \times & & \times & \times \\ \times & \times & \times & \end{smallmatrix}$ may be taken in the order 1324, but not in the order 1234). Thus, if $\alpha_1, \dots, \alpha_p$ are the numbers of non-zero elements in the 1st, ..., pth columns and τ is the total number of rows in the array, we have

$$\tau \le \alpha_1 + (\alpha_2 - 1) + \dots + (\alpha_p - 1) = \Sigma\alpha_h - p + 1. \tag{9}$$

Now if this array is one of those contributing to the coefficient of $\kappa_{t_1}\dots\kappa_{t_p}$ in $\kappa_{(2)}(r_1\dots r_p)$ (so that $\Sigma t_j = \Sigma r_h$) then the corresponding term is of order $\nu^{\Sigma(t_j-1)} = \nu^{-\Sigma r_h + \tau}$; for the numerical factors and pattern functions for each separation of the array do not depend upon ν, while each κ_{t_j} is of order ν^{t_j-1}. But (9) shows that

$$-\Sigma r_h + \tau \le -\Sigma(r_h - \alpha_h) - p + 1 \le -p + 1; \tag{10}$$

for α_h is the number of parts in a partition of r_h, and so does not exceed r_h. Thus this particular term of $\kappa_{(2)}(r_1\dots r_p)$ is of order ν^{-p+1} or lower. Hence the same is true of $\kappa_z(r_1\dots r_p)$ itself, and finally, by (5), of κ_{py}.

2. By the use of multivariate sampling rules more general results than Theorem 1 can be proved. Perhaps the most general is Theorem 2.

THEOREM 2. *If* $y^1 = f^1(x^1, \dots, x^p)$, $y^2 = f^2(x^1, \dots, x^p)$, ... *are functions of the variates* $x^1, \dots, x^p$, formally expansible in the forms*

$$y^h = c^h + \Sigma c_i^h x^i + \Sigma c_{ij}^h x^i x^j + \dots, \tag{11}$$

---

* The superfixes are indices, not exponents.

...and if the pth order cumulant... ...ts of order... ...for... ...then the same holds for the cumulants... ...of the q^k. Here we use a slightly modified form of the 'tensor' notation suggested by Kaplan (1952).)

An outline of the proof is as follows. If... ...denote the quantities $z^1, z^2, z^1z^1, z^1z^2, ..., z^az^b, z^1z^1z^1, ...$ written in some convenient order, then (11) may be rewritten in the form

$$y^A = \Sigma c^{Aa} z_a, \tag{12}$$

whence we easily derive

$$\kappa_y^{A_1...A_p} = \sum_{a_1,...,a_p} c^{A_1a_1}...c^{A_pa_p}\kappa_{(z)}(a_1...a_p), \tag{13}$$

where $\kappa_{(z)}(s_1...s_p)$ is the mixed cumulant of $z_{a_1},...,z_{a_p}$. Thus it suffices to show that all the $\kappa_{(z)}(s_1...s_p)$ are of order ν^{-p+1} or lower, for $p = 1, 2, ...$. The proof is completed in much the same way as before, but in order to see clearly the application of the results of my other paper the z's should be relabelled as follows: $z^i = x^i, z^{ij} = x^ix^j, ...$.

3. *An example.* Johnson & Welch (1939) have found by direct calculation, using Stirling's asymptotic expansion of the gamma function, that if χ^2 is the sum of squares of ν independent standard normal deviates, then the cumulants of χ, as far as the sixth, are given asymptotically as follows:

$$\begin{aligned}
\kappa_{1\chi} &\sim \nu^{\frac{1}{2}}, & \kappa_{2\chi} &\sim \tfrac{1}{2}, \\
\kappa_{3\chi} &\sim \tfrac{1}{4}\nu^{-\frac{1}{2}}, & \kappa_{4\chi} &\sim \tfrac{1}{16}\nu^{-2}, \\
\kappa_{5\chi} &\sim -\tfrac{1}{16}\nu^{-1}, & \kappa_{6\chi} &\sim -\tfrac{11}{32}\nu^{-3}.
\end{aligned} \tag{14}$$

Now if we write $x = \chi^2 \nu - 1$, $y = \sqrt{(1+x)} = \chi \div \sqrt{\nu}$, then $\kappa_{\rho x}$ is of order $\nu^{-\rho+1}$, so that Theorem 1 shows that $\kappa_{\rho y}$ is of order $\nu^{-\rho+1}$ or lower and $\kappa_{\rho\chi}$ is of order $\nu^{-\frac{1}{2}\rho+1}$ or lower. Thus the odd cumulants in (14) are of the order we should expect, but the even ones (apart from the second) seem to be a whole order lower than demanded by Theorem 1. I have been unable to prove that $\kappa_{2\rho,\chi} = O(\nu^{-\rho})$ for general values of ρ ($\geqslant 2$).

REFERENCES

JAMES, G. S. (1955). On moments and cumulants of systems of statistics. (Not yet published.)

JOHNSON, N. L. & WELCH, B. L. (1939). On the calculation of the cumulants of the χ-distribution. *Biometrika*, **31**, 216–18.

KAPLAN, E. L. (1952). Tensor notation and the sampling cumulants of k-statistics. *Biometrika*, **39**, 319–23.

KENDALL, M. G. (1947). *The Advanced Theory of Statistics*, vol. 1, 3rd ed. London: Charles Griffin and Co.

The likelihood ratio test for Markoff chains

BY I. J. GOOD

1. The present paper is virtually a footnote to that of Hoel (1954), who, using and acknowledging the methods of Bartlett (1951), constructed a likelihood ratio test for the order of a Markoff chain. If H_ν is the hypothesis that the chain is of order ν, then Hoel tests $H_{\nu-1}$ against (or rather *within*) H_ν. Here we make a simple generalization so as to test H_μ within H_ν, and we relate the work to some previous work.

2. Let $x_1, x_2, ..., x_N$ be a sequence, $\mathfrak{S}$, of observations, each observation being capable of taking one of t values denoted conventionally by $1, 2, ..., t$. Let H_ν be the hypothesis that $\mathfrak{S}$ is a Markoff chain of order ν ($\nu = 0, 1, 2, ...$). Note that H_0 means that $\mathfrak{S}$ is a random sequence. We may obtain uniformity in expression by defining H_{-1} to mean that $\mathfrak{S}$ is a 'perfectly' random or 'equiprobably' random sequence. H_{-1} is the only H_μ that is a simple statistical hypothesis. Clearly H_μ implies H_ν whenever $\mu < \nu$.

The likelihood ratio test for composite hypotheses may be expressed in the following form. Let H and H' be hypotheses such that H implies H', so that H' at least is not a simple statistical hypothesis. Let E be an experimental result and let

$$\lambda = \max P(E \mid H^*)/\max P(E \mid H'^*),$$

the maxima being taken over all simple statistical hypotheses H^*, H'^* belonging to H and H' respectively. λ is the likelihood ratio statistic for composite hypotheses and we may say that it tests H *within* H'. We may call H the null hypothesis whether or not it is simple and (within orthodox statistics) λ can be used for the rejection of H.

Hoel (1954) used λ for testing $H_{\nu-1}$ within H_ν ($\nu \geqslant 1$), and, with suitable conventions, his result applies also for $\nu = 0$. Bartlett (1951) used the likelihood ratio statistic for testing completely specified chains of order $\nu - 1$ within H_ν. Since H_{-1} can be regarded as a completely specified chain of *any* order, Bartlett's work included tests for H_{-1} within H_ν ($\nu = 0, 1, 2, \ldots$). Similarly, a completely specified chain of order μ is also one of order $\nu - 1$ (if $\mu < \nu$), so that Bartlett's work applies for testing any completely specified chain of order μ within H_ν, where $\mu < \nu$.

We shall here generalize Hoel's work in order to test H_μ within H_ν where $\mu < \nu$. Only when $\mu = -1$ does the generalization overlap with Bartlett's results, since the only H_μ that is completely specified is H_{-1}. (A completely specified statistical hypothesis is the same thing as a simple statistical hypothesis.)

3. Associated with $\mathfrak{S}$ is the corresponding cyclic sequence $\overline{\mathfrak{S}}$ defined by regarding the first element of $\mathfrak{S}$ as immediately following its last one. We shall always denote properties of $\overline{\mathfrak{S}}$ by placing a bar over the corresponding algebraic symbol relating to $\mathfrak{S}$.

A sequence of ν consecutive observations is called a ν-*sequence*. Let $n_{r_1, \ldots, r_\nu}$ or $n_\mathbf{r}$ for short, be the number of ν-sequences in $\mathfrak{S}$ which are $(r_1, r_2, \ldots, r_\nu) = \mathbf{r}$. Let $\bar{n}_\mathbf{r}$ be defined similarly. Let

$$\psi_\nu^2 = \sum_\mathbf{r} \left(n_\mathbf{r} - \frac{N-\nu+1}{\nu}\right)^2 \Big/ \frac{N-\nu+1}{\nu} \quad (\nu \geqslant 1), \tag{1}$$

$$\bar{\psi}_\nu^2 = \sum_\mathbf{r} (\bar{n}_\mathbf{r} - Nt^{-\nu})^2 / Nt^{-\nu} \quad (\nu \geqslant 1), \tag{2}$$

$$\psi_0^2 = \bar{\psi}_0^2 = 0, \tag{3}$$

$$K_\nu = 2 \sum_\mathbf{r} n_\mathbf{r} \log n_\mathbf{r} \quad (\nu \geqslant 1), \tag{4}$$

$$K_0 = 2(N-\nu+1)\log(N-\nu+1), \quad K_{-1} = 0, \tag{5}$$

$$\bar{K}_\nu = 2 \sum_\mathbf{r} \bar{n}_\mathbf{r} \log \bar{n}_\mathbf{r} \quad (\nu \geqslant 1), \tag{6}$$

$$\bar{K}_0 = 2N \log N, \quad \bar{K}_{-1} = 0, \tag{7}$$

$$\nabla \psi_\nu^2 = \psi_\nu^2 - \psi_{\nu-1}^2, \quad \text{etc.}, \tag{8}$$

$$\nabla^2 \psi_\nu^2 = \psi_\nu^2 - 2\psi_{\nu-1}^2 + \psi_{\nu-2}^2, \quad \text{etc.} \tag{9}$$

(The logarithms are to base e and $0 \log 0$ means 0.)

It is known that if H_{-1} is true, then ψ_ν^2 and $\bar{\psi}_\nu^2$ do not have asymptotically gamma variate (chi-squared) distributions, but $\nabla \psi_\nu^2$, $\nabla \bar{\psi}_\nu^2$, $\nabla^2 \psi_\nu^2$, and $\nabla^2 \bar{\psi}_\nu^2$ do (see Bartlett, 1951; Good, 1953). Moreover (precisely),

$$\mathcal{E}\psi_\nu^2 = \mathcal{E}\bar{\psi}_\nu^2 = t^\nu - 1, \tag{10}$$

while $\nabla \psi_\nu^2$ and $\nabla \bar{\psi}_\nu^2$ have ∇t^ν degrees of freedom ($\nu \geqslant 1$), and $\nabla^2 \psi_\nu^2$ and $\nabla^2 \bar{\psi}_\nu^2$ have $\nabla^2 t^\nu$ degrees of freedom ($\nu \geqslant 2$).

Hoel defined λ as the ratio of the maximum likelihood given $H_{\nu-1}$ to that given H_ν and obtained the asymptotic distribution of λ, given $H_{\nu-1}$. Let us denote by $\lambda_{\mu,\nu}$ the ratio of the maximum likelihood given H_μ to that given H_ν ($\mu < \nu$). Then Hoel's λ is our $\lambda_{\nu-1,\nu}$ and obviously $\lambda_{\mu,\nu}$ is the product of $\nu - \mu$ of Hoel's λ's, namely

$$\lambda_{\mu,\nu} = \lambda_{\mu,\mu+1} \lambda_{\mu+1,\mu+2} \cdots \lambda_{\nu-1,\nu}. \tag{11}$$

The expression given by Hoel for $-2 \log \lambda_{\nu-1,\nu}$ may be written in the form

$$-2 \log \lambda_{\nu-1,\nu} = \nabla^2 K_{\nu+1} \quad (\nu = 1, 2, 3, \ldots). \tag{12}$$

With our conventions this equation is also valid for $\nu = 0$.

Hoel's result can be stated in the form: When $H_{\nu-1}$ is given, $\nabla^2 K_{\nu+1}$ has asymptotically a gamma variate distribution with $\nabla^2 t^{\nu+1}$ degrees of freedom if $\nu \geqslant 1$ and we may add that for $\nu = 0$ it has asymptotically a gamma variate distribution with $t - 1$ degrees of freedom, since it then reduces to the likelihood ratio test for 'perfect' (equiprobable) randomness against randomness in general. Clearly

$$-2 \log \lambda_{\mu,\nu} = -2 \log \lambda_{\mu,\mu+1} - \ldots - 2 \log \lambda_{\nu-1,\nu}$$

$$= \nabla^2 K_{\nu+1} + \nabla^2 K_\nu + \ldots + \nabla^2 K_{\mu+2}$$

$$= \nabla K_{\nu+1} - \nabla K_{\mu+1} \quad (-1 \leqslant \mu < \nu). \tag{13}$$

If we could assume that the variables $-2\log \lambda_{\mu+1}$, $\nu = 0, 1, 2, \ldots$ were asymptotically independent, given H_p, it would follow that $-2\log \lambda_\mu$ has a gamma variate distribution with degrees of freedom

$$\begin{aligned}\nabla r^{\mu+1} - \nabla \nu^{\mu+1} \quad (\mu \geqslant 0), \\ \nabla r^{\mu+1} \quad (\mu = -1),\end{aligned} \tag{14}$$

and we could use these results for testing H_μ within H_p. Unfortunately, the independence does not seem to be easy to prove. Nevertheless, the result just stated is correct and can be proved by precisely the same method as Hoel used, except that his suffix r is to be replaced throughout his proof by a sequence of $r - \mu$ suffixes. It is unnecessary to repeat the argument since the modifications are entirely trivial. The conjecture that the variables $-2\log \lambda_{\mu+1}$ are asymptotically independent is strengthened by the knowledge that the above deduction from it is correct.

4. The value of the generalization of Hoel's results is that there may not be a significant distinction between 'adjacent' hypotheses in the sequence $H_0, H_1, \ldots, H_p$, yet H_μ may be clearly rejectable by the statistic $\nabla K_{\mu-1} - \nabla K_{\mu+1}$. If we put $\mu = 0$, $\mu = -1$ we have a test for randomness ('perfect' randomness) within Markovity of order r. If $\mu = 0$ and $r = 1$ we obtain a test that is a special case of the likelihood ratio test for contingency tables. (See, for example, Wilks, 1946, p. 220, where, however, there is a minor slip in that the expression given as λ is really $1/\lambda$.)

5. There is clearly a strong analogy between the expressions $\nabla K_{\mu-1} - \nabla K_{\mu+1}$ and $\nabla \psi_{\mu-1}^2 - \nabla \psi_{\mu+1}^2$. When H_{-1} is true the latter expression also has asymptotically a gamma variate distribution with the same number of degrees of freedom as the former expression, since (as shown by Good (1953)) the variables $\nabla^2 \psi_\nu^2$ $(\nu = 1, 2, \ldots)$ are asymptotically independent, at any rate when t is a prime number. This analogy is no coincidence since $\nabla \psi_\nu^2$ is the asymptotic form of $\nabla K_{\nu-1} - \nabla K_0$ when H_{-1} is true.

6. We may also use the cyclic definitions. When H_p is true, $\nabla K_{\mu-1} - \nabla K_{\mu+1}$ has asymptotically a gamma variate distribution with a number of degrees of freedom given by (14). The cyclic definition is mathematically simpler than the non-cyclic one and makes the checks

$$\sum_{r_p} n_r = n_{r_1, \ldots, r_{p-1}}, \quad \sum_{r_1} n_r = n_{r_2 \ldots r_p}$$

precise.

7. Not much is known concerning the accuracy of the asymptotic formulae when N is specified. When testing H_{-1} the psi-squared statistic has the advantage that its expected value is precisely known, but in principle this statistic may well be less powerful than the likelihood ratio. Unless N is very small there is probably little to choose between the cyclic and non-cyclic forms of the latter.

8. Another statistic that would be worth consideration would be $\nabla L_{\nu-1} - \nabla L_{\mu+1}$ (and its cyclic form), where

$$L_\nu = 2 \sum_r \log n_r! \quad (\nu \geqslant 1), \qquad L_0 = 2\log N!, \qquad L_{-1} = 0. \tag{15}$$

When $\nu = 1$ and $\mu = 0$ this statistic reduces to $\nabla^2 L_2$, which is minus twice the log likelihood of the $n_{r_\nu r_1}$'s, regarded as forming the interior of a contingency table for which the marginal totals are assigned and independence is assumed.

REFERENCES

BARTLETT, M. S. (1951). The frequency goodness of fit for probability chains. *Proc. Camb. Phil. Soc.* **47**, 86–95.

GOOD, I. J. (1953). The serial test for sampling numbers and other tests for randomness. *Proc. Camb. Phil. Soc.* **49**, 276–84.

HOEL, P. G. (1954). A test for Markoff chains. *Biometrika*, **41**, 430–3.

WILKS, S. S. (1946). *Mathematical Statistics*. Princeton.

Exact forms of some invariants for distributions admitting sufficient statistics

By V. S. HUZURBAZAR

University of Poona, India

1. INTRODUCTION

The concept of distance in statistics is due to Mahalanobis (1936) who defined the distance between two multivariate normal populations. Later Jeffreys (1946, 1948) defined a wider class of invariants of probability distributions which may be regarded as providing measures of distance between two

probability distributions, in general. If p and p' are the density functions of two probability distributions of a variate x, then the expressions defined by Jeffreys

$$I_m = \int | \, p'^{1/m} - p^{1/m} \, |^m \, dx, \tag{1}$$

and

$$J = \int (p'-p) \log \frac{p'}{p} \, dx, \tag{2}$$

are all positive definite and invariant for all non-singular transformations of the variate and the parameters. These expressions, therefore, provide measures of distance between two distributions. For discrete distributions p and p' are simply the probabilities that x takes a particular value r. Extension to multivariate distributions follows immediately.

An important case is the distance between two distributions having the *same mathematical form* but with *different* sets of values of the parameters. If the corresponding parameters in the two sets differ by infinitesimals, we get the differential forms of the invariants I_m and J. The invariants I_2 and J have interesting mathematical properties and have been used by Jeffreys in stating the prior probabilities of parameters in his theory of estimation and tests of significance. Though these invariants have not yet attracted the attention of the frequency theorists of probability, it is likely that they may find applications in their work also.

Jeffreys has obtained the exact forms of the invariants I_2 and J for the univariate and bivariate normal distributions, the Poisson and the binomial distributions; and for these distributions the exact forms come out as *explicit* functions of the parameters of distributions.

The object of this paper is to prove, in general terms, a remarkable property that for all distributions admitting sufficient statistics the exact forms of I_m (m even) and J come out as *explicit* functions of the parameters of distributions. It may be mentioned here that the properties of I_m ($m \neq 2$) have so far remained uninvestigated, the form of I_m being then very complicated.

2. Distributions admitting sufficient statistics

We shall take the most general form of distributions admitting sufficient statistics as given by Koopman (1936),

$$f(x, \alpha_j) = \exp\left\{ \sum_{k=1}^{p} u_k(\alpha_j) v_k(x) + A(x) + B(\alpha_j) \right\}, \tag{3}$$

where, for brevity, (α_j) denotes the set of p parameters $(\alpha_1, \alpha_2, ..., \alpha_p)$; u_k and B are functions of the α_j, and v_k and A are functions of x. For multivariate distributions (x) is to be replaced by the set of variates $(x_1, x_2, ..., x_q)$.

Since $\int f(x, \alpha_j) \, dx \equiv 1$, for all α_j, we have

$$\int \exp\left\{ \sum_{k=1}^{p} u_k(\alpha_j) v_k(x) + A(x) \right\} dx = \exp\{- B(\alpha_j)\}. \tag{4}$$

Now the $u_k(\alpha_j)$ are p independent functions of the p parameters α_j. We can express the α_j inversely as functions of the u_k's. Then $B(\alpha_j)$ can be expressed in terms of the u_k's as

$$B(\alpha_j) = b(u_k). \tag{5}$$

Then (4) becomes

$$\int \exp\left\{ \sum_{k=1}^{p} u_k(\alpha_j) v_k(x) + A(x) \right\} dx = \exp\{- b(u_k)\}. \tag{6}$$

3. Exact form of I_2

Let

$$f(x, \alpha_j) = \exp\left\{ \sum_{k=1}^{p} u_k(\alpha_j) v_k(x) + A(x) + B(\alpha_j) \right\}, \tag{7}$$

and

$$f(x, \alpha_j') = \exp\left\{ \sum_{k=1}^{p} u_k(\alpha_j') v_k(x) + A(x) + B(\alpha_j') \right\}. \tag{8}$$

so that $f(x, \alpha_j)$ and $f(x, \alpha'_j)$ have the same mathematical form but different sets of values of the parameters. We have

$$I_2 = \int \{\sqrt{[f(x, \alpha_j)]} - [f(x, \alpha'_j)]\}^2 \, dx$$

$$= 2 - 2 \int \{f(x, \alpha_j) f(x, \alpha'_j)\}^{\frac{1}{2}} \, dx$$

$$= 2 - 2 \int \exp\left[\sum_{k=1}^{p} \{\tfrac{1}{2} u_k(\alpha_j) + \tfrac{1}{2} u_k(\alpha'_j)\} v_k(x) + A(x) + \tfrac{1}{2} B(\alpha_j) + \tfrac{1}{2} B(\alpha'_j) \right] dx$$

$$= 2 - 2 \exp\{\tfrac{1}{2} B(\alpha_j) + \tfrac{1}{2} B(\alpha'_j)\} \int \exp\left[\sum_{k=1}^{p} \{\tfrac{1}{2} u_k(\alpha_j) + \tfrac{1}{2} u_k(\alpha'_j)\} v_k(x) + A(x) \right] dx. \tag{9}$$

Writing $\tfrac{1}{2} u_k(\alpha_j) + \tfrac{1}{2} u_k(\alpha'_j)$ for $u_k(\alpha_j)$ in (6), we have

$$\int \exp\left[\sum_{k=1}^{p} \{\tfrac{1}{2} u_k(\alpha_j) + \tfrac{1}{2} u_k(\alpha'_j)\} v_k(x) + A(x) \right] dx = \exp[-b\{\tfrac{1}{2} u_k(\alpha_j) + \tfrac{1}{2} u_k(\alpha'_j)\}]. \tag{10}$$

From (9) and (10) we have

$$I_2 = 2 - 2 \exp[\tfrac{1}{2} B(\alpha_j) + \tfrac{1}{2} B(\alpha'_j) - b\{\tfrac{1}{2} u_k(\alpha_j) + \tfrac{1}{2} u_k(\alpha'_j)\}]. \tag{11}$$

The curious point to be noted is that the function $A(x)$ remains unaltered in the integral on the left-hand side of (10) which enables us to evaluate that integral explicitly in virtue of (6).

3·1. *Illustrative example*

Consider the Type III distribution

$$f(x, a, p) = \frac{a^p e^{-ax} x^{p-1}}{\Gamma(p)} \quad (0 \leqslant x < \infty). \tag{12}$$

We write
$$f(x, a, p) = \exp\{-ax + p \log x - \log x + p \log a - \log \Gamma(p)\}. \tag{13}$$

Here $u_1 = a$, $u_2 = p$, $B = p \log a - \log \Gamma(p)$. Expressing B in terms of u_1 and u_2,

$$B = u_2 \log u_1 - \log \Gamma(u_2) = b(u_1, u_2),$$

$$b\left(\frac{a+a'}{2}, \frac{p+p'}{2}\right) = \left(\frac{p+p'}{2}\right) \log \frac{a+a'}{2} - \log \Gamma\left(\frac{p+p'}{2}\right).$$

$$I_2 = 2 - 2 \exp\left[\tfrac{1}{2} p \log a - \tfrac{1}{2} \log \Gamma(p) + \tfrac{1}{2} p' \log a' - \tfrac{1}{2} \log \Gamma(p') - \left(\frac{p+p'}{2}\right) \log\left(\frac{a+a'}{2}\right) + \log \Gamma\left(\frac{p+p'}{2}\right) \right]$$

$$= 2 - \frac{2 \Gamma\left(\dfrac{p+p'}{2}\right)}{\{\Gamma(p) \Gamma(p')\}^{\frac{1}{2}}} \frac{a^{\frac{1}{2}p} a'^{\frac{1}{2}p'}}{\left(\dfrac{a+a'}{2}\right)^{\frac{1}{2}(p+p')}}. \tag{14}$$

4. THE EXACT FORM OF J

$$J = \int \{f(x, \alpha'_j) - f(x, \alpha_j)\} \{\log f(x, \alpha'_j) - \log f(x, \alpha_j)\} \, dx$$

$$= \sum_{k=1}^{p} \{u_k(\alpha'_j) - u_k(\alpha_j)\} \left\{ \int v_k(x) f(x, \alpha'_j) \, dx - \int v_k(x) f(x, \alpha_j) \, dx \right\}$$

$$= \sum_{k=1}^{p} \{u_k(\alpha'_j) - u_k(\alpha_j)\} \{E' v_k(x) - E v_k(x)\}, \tag{15}$$

where $E' v_k(x)$ and $E v_k(x)$ denote the expectations of $v_k(x)$ when the parameters are α'_j and α_j respectively. For brevity write

$$E'_k = E' v_k(x), \quad E_k = E v_k(x).$$

Then
$$J = \sum_{k=1}^{p} \{u_k(\alpha'_j) - u_k(\alpha_j)\} (E'_k - E_k). \tag{16}$$

E'_k and E_k can be obtained as follows. We have

$$\frac{\partial}{\partial \alpha_r} \log f(x, \alpha_j) = \sum_{k=1}^{p} \frac{\partial u_k}{\partial \alpha_r} v_k(x) + \frac{\partial B}{\partial \alpha_r}.$$

Since

$$E\left\{\frac{\partial}{\partial \alpha_r} \log f(x, \alpha_j)\right\} \equiv 0,$$

we have

$$\sum_{k=1}^{p} \frac{\partial u_k}{\partial \alpha_r} E_k + \frac{\partial B}{\partial \alpha_r} = 0. \tag{17}$$

Setting $r = 1, 2, \ldots, p$ in (17), we have p simultaneous linear equations to determine the E_k. Then E'_k are obtained by writing α'_j for α_j in E_k. Thus (15) and (17) enable us to express J explicitly in terms of the parameters.

The above formulae are greatly simplified if we take the u_k as parameters and express $B(\alpha_j)$ in terms of them. Then (17) becomes simply

$$E_r + \frac{\partial B}{\partial u_r} = 0, \quad \text{so that} \quad E_r = -\frac{\partial B}{\partial u_r}.$$

From (16),

$$J = \sum_{k=1}^{p} (u'_k - u_k)\left(-\frac{\partial B'}{\partial u'_k} + \frac{\partial B}{\partial u_k}\right). \tag{18}$$

Writing $u'_k - u_k = du_k$, the differential form of J is

$$J = -\sum_{k=1}^{p} \sum_{l=1}^{p} \frac{\partial^2 B}{\partial u_k \partial u_l} du_k du_l. \tag{19}$$

5. THE EXACT FORM OF I_m (m EVEN)

For brevity write

$$f(x, \alpha_j) = f, \quad f(x, \alpha'_j) = f', \quad u_k(\alpha'_j) = u'_k, \quad B(\alpha'_j) = B', \quad \text{etc.}$$

Now

$$I_m = \int |f'^{1/m} - f^{1/m}|^m dx$$

$$= \int (f'^{1/m} - f^{1/m})^m dx \quad \text{(since m is even)}$$

$$= \int \left\{\sum_{r=0}^{m} (-1)^r \binom{m}{r} f'^{(m-r)/m} f^{r/m}\right\} dx$$

$$= \sum_{r=0}^{m} (-1)^r \binom{m}{r} \lambda_r, \tag{20}$$

where

$$\lambda_r = \int f'^{(m-r)/m} f^{r/m} dx$$

$$= \int \exp\left\{\sum_{k=1}^{p}\left(\frac{m-r}{m} u'_k + \frac{r}{m} u_k\right) v_k(x) + A(x) + \frac{m-r}{m} B' + \frac{r}{m} B\right\} dx$$

$$= \exp\left(\frac{m-r}{m} B' + \frac{r}{m} B\right) \int \exp\left\{\sum_{k=1}^{p}\left(\frac{m-r}{m} u'_k + \frac{r}{m} u_k\right) v_k(x) + A(x)\right\} dx. \tag{21}$$

Writing $\frac{m-r}{m} u'_k + \frac{r}{m} u_k$ for u_k in (6), we have

$$\int \exp\left\{\sum_{k=1}^{p}\left(\frac{m-r}{m} u'_k + \frac{r}{m} u_k\right) v_k(x) + A(x)\right\} dx = \exp\left\{-b\left(\frac{m-r}{m} u'_k + \frac{r}{m} u_k\right)\right\}. \tag{22}$$

From (21) and (22),

$$\lambda_r = \exp\left\{\frac{m-r}{m} B' + \frac{r}{m} B - b\left(\frac{m-r}{m} u'_k + \frac{r}{m} u_k\right)\right\}.$$

Then from (20), $I_m = \sum_{r=0}^{m} (-1)^r \binom{m}{r} \exp\left\{\frac{m-r}{m} B' + \frac{r}{m} B - b\left(\frac{m-r}{m} u_a' + \frac{r}{m} u_b\right)\right\}.$ (23)

Curiously again, the function $A(c)$ remains unaltered in the integral in (22), which enables us to evaluate that integral explicitly in virtue of (6).

REFERENCES

JEFFREYS, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc.* A, **186**, 453.

JEFFREYS, H. (1948). *Theory of Probability*. 2nd ed. Oxford: Clarendon Press.

KOOPMAN, B. O. (1936). On distributions admitting a sufficient statistic. *Trans. Amer. Math. Soc.* **39**, 399.

MAHALANOBIS, P. C. (1936). On the generalized distance in statistics. *Proc. Nat. Inst. Sci. Ind.* **12**, 49.

REVIEWS

Demand Analysis. A Study in Econometrics. By HERMAN WOLD, in association with LARS JURÉEN. New York: Wiley and Sons (Stockholm: Almqvist and Wiksell). Pp. 1953. xvi + 358. 56s.

Originating in a study of consumer demand for food in Sweden undertaken in 1938 by Prof. Wold and Mr Jureen, this book sets out to present a self-contained account of both methods and results. It includes a report on the empirical findings, based on family budget surveys and market statistics, a section giving a survey of methods in non-technical language, and sections dealing in more concentrated and technical manner with some aspects of economic theory, stochastic processes and regression analysis.

The theoretical sections are of considerable incidental interest for econometrics generally, although the developments, on an abstract level, are occasionally only remotely relevant to the reported empirical study.

The author states that one of his principal aims was to justify the use of the traditional method of least-squares regression, via the use of economic models of so-called recursive type. His main argument rests on the notion of causality in consumer demand situations. With this as a guide to the method of application, he obtains, among others, the result that the regression method for the estimation of recursive systems is unbiased and consistent on the assumption of non-correlation between the explanatory variables and the disturbances, irrespective of whether there is inter-correlation between these disturbances. A generalization of the classical formula is derived for the standard error of a regression coefficient which allows for intercorrelation.

The wide scope of recursive models is indicated by a proof that any set of time series has a formal representation in terms of a recursive system.

However the causality argument is regarded (and it is certainly not a straightforward one), the regression method would still appear on Wold's showing to compare favourably, in demand analysis at least, with other methods which have been proposed.

The book is very fully annotated; there is an extensive bibliography (although several of the dated references in the text are omitted), and important sets of exercises accompany the theoretical parts. Based on years of research into the foundations of the subject, this is a very thorough work of scholarship.

F. G. FOSTER

A Study in the Analysis of Stationary Time-series. 2nd edition. By HERMAN WOLD, with an Appendix by PETER WHITTLE. Stockholm: Almqvist and Wiksell. 1954. Pp. viii + 236. 42s.

As far as the main text is concerned, the present edition differs little from the first, published in 1938, the only change of major importance being the replacement of appendices A and B in the first edition by two new ones.

In the first appendix, Prof. Wold makes short but useful comments on certain parts of the text, mainly in the light of recent work. Appendix 2, comprising 32 pages, is written by Dr P. Whittle and is devoted to a survey of recent advances in the subject.

The first three chapters of the book dealing with Prof. Wold's fundamental contributions to the theory of discrete stationary schemes are by now familiar to most statisticians and need no further comment. Chapter IV, on the application of stationary schemes to experimental data, has not been as successful in withstanding the passage of time, and the discussion tends to be outmoded in view of the advances made in the testing of specific hypotheses and in the problem of estimation.

Chapter 1 of Appendix 2 is devoted to the contributions of Cramer, Kolmogoroff and Wiener to spectral theory and Chapter 2 to a synopsis of Whittle's own work in the theory of inference, together with some miscellaneous results in distribution theory and periodogram analysis. To attempt even a cursory review of the field in so short a space is not practicable, and the result is a one-sided exposition of the subject. Even so, the treatment is interesting and some new results are introduced.

The emphasis laid throughout on the mathematical structural, rather than the statistical, aspects of time series analysis, is such that the book is of greater value to the mathematical statistician than to one who is interested in the approaches to economic process methodology, etc. The latter is also likely to view with extreme pessimism the various kinds of intervals obtained, the lack about the probability of the information provided by the analyses. Broadly stated, the argument is that significance problems are complicated by the fact that one is determining the whole structure of a series in one a realization or sample and also that the estimation of quantities of significance to the sample functions is not possible since they are conditioned by the size of the statistical masses to which they refer, in particular, there is arbitrariness in (a) the time unit of collection of the data, (b) the region in space to which this unit refers. A great deal has been done to remove the first objection by realizing the futility of testing individual serial correlations, and much progress has been made in discriminating between various hypotheses. The evidence up to the present seems to indicate that the problem is soluble in terms of the classical theory.

The second objection is more troublesome but is not so serious as one might be led to expect. As far as (a) is concerned, the choice of time unit raises no special difficulty, since in many instances the auto-correlation properties of time series are not in themselves of fundamental importance. Thus whilst collecting data at quarterly intervals instead of annually is likely to alter the structural form of the series, if it were required to estimate the correlation between two time series, the ultimate aim would be to eliminate the effect of autocorrelation during the course of the analysis. The intrinsic autocorrelative properties of a series are likely to be more important in the problem of prediction, but recent work seems to suggest that as far as practical applications are concerned much remains to be done in this field.

The difficulties raised by (b) are more serious, but this type of objection is latent in, and tends to restrict the conclusions which are drawn in many other problems in statistics.

The need for further research is obvious, and Prof. Wold's own work on this problem of quantitative significance is sufficient to raise the question that perhaps our assumptions in dealing with observed series are too stringent.

G. M. JENKINS

An Introduction to Stochastic Processes with special reference to Methods and Applications. By M. S. BARTLETT. Cambridge University Press. 1955. Pp. xiv + 312. 35s.

Probability problems concerned with the description of changes with time —where we use 'time' in the operational sense—are as old as the theory itself. The entire development of the subject in the sixteenth and seventeenth centuries was due to the zeal and passion of gamblers and, among other queries, the probability of ruin was one which ranked importantly. Interest in this type of problem never faded out, and we find it discussed by one after another of the great probabilists right up to the present day. The difference between the moderns and their predecessors is that problems involving the dynamics of change are now given a special title —stochastic processes – and are subject to a unified systematic attack.

This systematic attack, which has its origins chiefly in the research of the Russian school of probability and in particular in that of Kolmogoroff, has resulted in a great deal of elegant mathematics and a certain amount of useful statistical methods. Its great defect has been that the problem of stochastic processes has been lifted beyond the understanding of many statisticians who still continue to solve the problems in *ad hoc* fashion as they arise.

In 1946 M. S. Bartlett attempted to remedy this state of affairs by publishing in mimeographed form the lectures which he had given in the University of North Carolina. Further clarification, in the discrete case, came with W. Feller's book *Introduction to Probability Theory*, and now we again have Prof. Bartlett who attempts to avoid too much mathematical abstraction and gives a general survey of techniques and applications. Topics on pure theory covered by him are random sequences, processes in continuous time, limiting stochastic operations and stationary processes. The problem of the random walk, Markoff chains, renewal processes, stochastic convergence and spectral analysis are included. Under statistical applications we have a chapter on applications of the random walk, queues, population growth and epidemic models, a chapter on statistical inference in stochastic processes and a chapter on the correlation analysis of time series including harmonic analysis.

This book is the best which has yet been produced on this specialized topic. It is lucidly written by someone who clearly sees the statistical implications of the abstract theory and may be read without undue difficulty by any student who has two years of university mathematical training behind him. It will undoubtedly be found indispensable by anyone wishing to learn something of a subject which has engaged the attention of nearly all probabilists in the past decade.

F. N. DAVID

Probability Theory. By M. Loève. New York: D. van Nostrand Company, Inc. 1955. Pp. xv + 501. $12.00.

This is a book on probability theory written by a mathematician who has himself made distinguished contributions to the subject during the past twenty years. Prof. Loève divides his book into five parts. In the introductory part we have what might be called the intuitive approach to probability, which is very often all that the statistician needs. In spite of the fact that the topics are not new they are set out in clear and pleasing fashion, and the corollaries to the various theorems proved, particularly in the laws of large numbers, will be fresh to those who are not thoroughly familiar with the French and Russian literature. The notion of chain dependence is also introduced at this stage. All this is introductory.

In Part One proper Prof. Loève sets us sternly to work with 83 pages on 'Notions of Measure Theory'. Here we have additive set functions, topological spaces, measurable functions and Lesbesgue integration. The two chapters which compose this part can be read fairly easily by a student who has taken a university degree course in mathematics, but the reasoning is rather condensed and may cause difficulty to students coming to it for the first time. In Part II, entitled 'General Concepts and Tools of Probability Theory', we have the development of probability laws, characteristic functions and distribution functions, much of which is in essence familiar although not perhaps the rigorous presentation.

With 'Independence', which is the topic of Part Three, Prof. Loève gives us a characteristically thorough thrashing out of convergence in probability and central limit theorems. Little of this has appeared explicitly in English, and it is useful to have it all down in a moderately compact form (107 pages). From 'Independence' it is natural that we should turn to 'Dependence', the topic of Part Four. Here, apart from the clear and mathematically rigorous presentation, there is little which has not been covered by Doob and Bartlett, except perhaps the second-order random functions.

This book will be excellent for those with a mathematical training who wish to specialize in probability theory regarded as a mathematical discipline. It is throughout logically inevitable in its presentation and written as rigorously as only a French-trained mathematician can. There can be no better beginning for a young mathematical student than to be given this book. It will, however, not be helpful in the training of the statistics student. Many a promising statistician has been wrecked on the desert island of mathematical rigour, and those trained in modern mathematical techniques often find the necessarily heuristic approach to numerical data painful. However, for the class of student indicated the book is excellent and all providing. It is unfortunate that it has to be recorded that the price is $12.

<div align="right">F. N. DAVID</div>

Statistical Inference. By Helen M. Walker and Joseph Lev. New York: Henry Holt and Company. 1953. Pp. xi + 510. 48s.

This text-book on statistical method is designed for beginners with little or no mathematical background. A good many mathematical formulae are, however, used (albeit always accompanied by careful explanation and practice exercises), and a student who has mastered such a text as Prof. Walker's previous book, *Mathematics Essential for Elementary Statistics*, would find the going easier. In choice of subject-matter and general emphasis it is definitely 'applied', the application being to educational research; this is not, however, revealed by the title.

In scope it is fairly comprehensive. It begins with the general idea of statistical inference and an intuitive discussion of hypothesis testing. The concept of probability is introduced, then the binomial distribution, with a discussion of the closeness of the normal approximation. By use of a population of two classes only, a variety of problems of hypothesis testing and estimation are treated, the account including such topics as one- and two-sided tests, critical regions, errors of the first and second types, power functions, confidence limits — concepts which gain in clarity for the beginner by definition in this simple context. Populations with more than two classes are next studied, and then some of the more important concepts regarding distributions on a continuous variable. There follows an account of a class project in sampling which is designed to familiarize the reader with the statistics possessing one of the four continuous distributions of major importance: the Normal, Chi-square, Student's and F-distributions. The method here is to plot the empirically derived distribution, superimpose the mathematical curve, and then introduce the student to the table of this function in place of its mathematical formula. With this basic material assembled, the subjects now treated include: inferences concerning

the mean or difference between two means, differences comparing variances and standard deviations, analysis of variance, linear regression and correlation, biserial, point biserial, tetrachoric and rank order correlation coefficients and the correlation ratio, the effects of measurement errors and the reliability coefficient, multiple regression and correlation, analysis of variance with two or more variables of classification, analysis of covariance, percentiles, standard scores and scales, non-parametric methods. The chapter on non-parametric methods was written by Prof. Lincoln Moses.

The treatment is concrete and as far as possible basic ideas are explained by means of numerical examples. The student is aided in computing techniques, a rather extensive set of twenty-three tables is an integral part of the text for this purpose, detailed instructions in their use being given. The reader may on occasion be somewhat startled at being addressed directly from the text, exhorted to do such and such a computation. A welcome feature of the book to correlationists should be the amount of illustrative numerical data actually taken from published work of educational research.

The book is well referenced, and a list of authors for further reading is placed at the ends of chapters. Many exercises are set, with answers given at the back of the book. There is a subject index, an author index and a glossary of the mathematical symbols used.

It is to be expected of a text with which Prof. Walker is associated that the exposition should be pedagogically sound, and there is evidence that a great deal of care was taken about the order and method of presentation of the material. The result cannot be said to be completely satisfactory. It is, for example, open to serious doubt on occasion whether the average student will be able to absorb simultaneously details of computing techniques and the basic principle of the statistical method under explanation. Occasionally, too, the style becomes so woolly as to present quite inaccurate information. Thus, on p. 14, two observations are defined to be independent 'when information about one of them provides no clue whatever as to the other'. A reader may well be puzzled by this if he considers that in the case of two independent observations from the same (unknown) population, the first observation provides an unbiased estimate of the mean of the second.

Despite some defects, however, research students of education and allied fields should find this text invaluable, not only as an instructional manual, but also as a general work of reference in statistical method.

F. G. FOSTER

Outline of Biometry and Statistics.. By C. I. BLISS and D. W. CALHOUN. New Haven, Conn.: Yale Co-operative Corporation. 1954. Pp. 272 + xvi. $4.50.

This book will be extremely useful to those who teach biometry, and to statisticians who work in a biological field or who wish to have biological examples to illustrate their teaching of statistics. Each chapter consists of a number of very concise explanatory notes, together with several illustrative worked examples and exercises chosen from a wide range of experimental material.

The subjects covered include the binomial, negative binomial and Poisson distributions; chi-squared and the analysis of contingency tables; the normal distribution with the estimation of its mean and variance; the analysis of variance, factorial experiments and simple experimental designs; regression, correlation and association; bio-assay. The authors regret that the proposed length of course (about 200 hours of students' time) necessarily excludes a few other topics such as covariance, probits, discriminants, and sequential and other sampling methods.

The professional statistician will find this book packed with information, such as short-cut formulas, precise statements of the limits within which approximate tests can be relied upon, and alternative parametric and non-parametric tests. A number of useful tables and charts are included. A long list of references enables the reader to consult the original papers whenever necessary, except for work yet unpublished. The exposition is almost always extremely lucid, and the reviewer has noticed few errors (apart from those already covered by a list of errata). But surely the exact test for 2×2 tables was independently discovered by Fisher and by Irwin (see *Metron*, **12**, 1935, 1)?

The book is too condensed for the unassisted student, but one working through the course under a teacher's guidance will find it very useful.

CEDRIC A. B. SMITH

Statistics and Mathematics in Biology. Edited by O. Kempthorne, T. A. Bancroft, J. W. Gowen and J. L. Lush. Ames, Iowa: Iowa State College Press. 1954. Pp. ix + 632. $6.75.

In case the title should mislead, it should be made clear at once that this is not a text-book on the use of statistical and mathematical methods in biology. It is a collection of forty-four essays by various authors (many of great distinction) on a wide range of topics in which biology and statistics meet. In many cases these deal with recent research by the author; some are highly theoretical, others concern practical problems involving only simple mathematics. Almost all are lucidly expressed, and can be read with interest and profit. The subjects discussed include causation, regression, path coefficients and multivariate analysis, classification problems, experimental designs, competition between species, growth curves, sampling of populations, toxicity tests, bio-assay, taste testing, feeding experiments, animal behaviour, gene frequency estimation, the effect of radiation on cells and viruses, genetic and breeding problems and the estimation of nucleoproteins in cell division.

Naturally this is not the sort of book to which one can turn for complete and detailed information on any one topic, although the very full bibliography may indicate the source of such information. But it is one which it would be useful to have about the laboratory (whether statistical or biological), since somewhere it may suggest the appropriate line of approach to a particular problem, or indicate a profitable new line of research.

One minor point may be mentioned. K. R. Nair has pointed out (*Bull. Calcutta Statist. Ass.* **20**, 1954, 18) that Quenouille's 'almost balanced incomplete block designs' are particular cases of the Bose-Nair-Rao 'partially balanced incomplete blocks' (Bose & Nair, *Sankhya*, **4**, 1939, 337).

CEDRIC A. B. SMITH

Statistical Analysis in Chemistry and the Chemical Industry. By C. A. Bennett and N. L. Franklin. New York: John Wiley and Sons, Inc. 1954. Pp. xvi + 724. 58s.

This book contains a well-written account of statistical techniques, some of which are not readily available elsewhere. After an introductory chapter, general notions of frequency distributions, measures of location, spread and association are introduced. There follows a discussion of probability theory including some account of moment generating functions, cumulants and k statistics and the common sampling distributions derived from the normal law. A short account of the properties of gamma and beta functions is followed by a chapter on confidence limits and tests of significance, including some non-parametric methods and a somewhat brief discussion of sequential tests. In Chapter 6 there is an account of linear regression, curvilinear regression and discriminant functions and of methods for the solution of the linear equations which arise in the application of these techniques.

Chapter 7, which contains over 150 pages, is concerned with the analysis of variance, and this is followed by a discussion of almost equal length on the design of experiments. The book ends with chapters on the analysis of counted data, quality control and tests for randomness.

To those workers in the chemical industry and elsewhere who have been introduced to statistical methods via a 'cookery book' and who now wish to broaden their knowledge and to learn something more of the basis of these methods, this book is recommended. It will be less useful to those who wish to know how statistics should be used in practice. Some of the examples fail to demonstrate the value of the method discussed, or the sort of circumstances in which it might be used, and instances occur where the statistical analysis confuses rather than clarifies the situation. This is particularly true of the example on p. 395, where the situation is much more readily appreciated from a set of four simple graphs than from the rather complicated analysis of variance which is given in this book.

It is very important that the experimenter should be the master and not the slave of statistical methods. In particular, he must be very clear about what it is he really wants to know and should not allow himself to be conditioned into using statistical ideas and concepts which are not really appropriate to his investigation. In particular, he should not allow himself to be overawed by apparently complicated mathematical machinery. When he feels his problem is not really answered by the application of standard techniques he should try to deal with it from first principles. His solution even if it is somewhat approximate and unorthodox will usually serve him better than the misapplication of a mathematically 'exact' technique. In helping to instil these first principles this text book is valuable; the reader should, however, not take too seriously some of the applications which are described.

G. E. P. BOX

A Million Random Digits with 100,000 Normal Deviates. The RAND CORPORATION Illinois. U.S.A.: The Free Press. 1955. Pp. xxv + 600. $10.00.

Since L. H. C. Tippett's set of 40,000 random numbers was first published in 1927 as *Tracts for Computers*, No. xv, there has been a steady demand for random numbers from a wide variety of users. This was recognized by the publication of M. G. Kendall's and B. Babington Smith's set of 100,000 digits in *Tracts for Computers*, No. xxiv, just prior to the war. Since then the demand has grown, rather than diminished, and the recent emphasis on Monte Carlo methods has led to the need for an even larger supply. This need has now been filled by the publication of this volume of 1,000,000 random digits produced with the aid of an electronic roulette wheel by the Rand Corporation. Tests for randomness, such as the frequency test, the so-called poker test and the pairs test, have been applied to the numbers and give satisfactory results.

Half the table has been used to obtain 100,000 random normal deviates by the conversion of five-digit random numbers with a table of the cumulative normal distribution function. The deviates obtained are tabulated to three decimal places.

<div align="right">P. G. MOORE</div>

Tablitsi dlya vichisleniya nepolnoi Γ-funktsii veroyatnosti χ^2 (Tables for the calculation of the incomplete Γ-function and the χ^2-probability function). By E. E. SLUTSKII (Ed. A. N. KOLMOGOROV). Moscow and Leningrad: Izdatelstvo Akademii Nauk SSSR. 1950. Pp. 14 + 55 pp. tables.

These tables are intended to facilitate the computation of

$$P(\chi^2, n) = \frac{1}{2^{\frac{1}{2}n-1}\Gamma(\frac{1}{2}n)} \int_{\chi}^{\infty} x^{n-1}e^{-\frac{1}{2}x^2}dx,$$

and the related incomplete gamma-function. There is a table of $P(\chi^2,n)$ for

$$\chi^2:\ 0\ (0\cdot1)\ 3\cdot2 \quad \text{and} \quad \chi^2:\ 3\cdot2\ (0\cdot2)\ 7\cdot0\ (0\cdot5)\ 10\cdot0\ (1\cdot0)\ 35\cdot0,$$
$$n:\ 0\ (0\cdot05)\ 0\cdot2\ (0\cdot1)\ 6\cdot0 \qquad\quad n:\ 0\ (0\cdot1)\ 0\cdot4\ (0\cdot2)\ 6\cdot0.$$

There are also tables of three auxiliary functions designed for special purposes. To facilitate interpolation for small values of χ^2 the function

$$T(\chi^2, n) = (\tfrac{1}{2}\chi^2)^{-\frac{1}{2}n}(1 - P(\chi^2, n))$$

is tabled for

$$\chi^2:\ 0\ (0\cdot05)\ 0\cdot2\ (0\cdot1)\ 1\cdot0,$$
$$n:\ 0\ (0\cdot05)\ 0\cdot2\ (0\cdot1)\ 6\cdot0.$$

For larger values of n the function $P(\chi^2, n)$ is tabled, with argument

$$t = \sqrt{(2\chi^2)} - \sqrt{(2n)}$$

in place of χ^2, for

$$t:\ -4\cdot0\ (0\cdot1)\ 4\cdot8,$$
$$n:\ 6\cdot0\ (0\cdot5)\ 11\cdot0\ (1\cdot0)\ 32\cdot0.$$

A further table gives values of the function $\pi(t, x) = P(\chi^2, n)$, where t is as above and $x = \sqrt{(2/n)}$. This function is tabled for

$$t = -4\cdot5\ (0\cdot1)\ 4\cdot8,$$
$$x = 0\ (0\cdot02)\ 0\cdot22\ (0\cdot01)\ 0\cdot25.$$

It is claimed that this small (seven page) table effectively provides for the computation of the incomplete gamma-function in the region ($n > 102$) not covered by the existing tables of K. Pearson.

A final table contains coefficients of Everett's and Newton's interpolation formulae. Second and fourth central differences with respect to χ^2 (or t), and second central differences with respect to n (or x) are printed in the tables.

The tables are clearly reproduced and there is an introduction setting out their genesis, calculation and use, with a number of worked examples.

<div align="right">N. L. JOHNSON</div>

OTHER BOOKS RECEIVED

1. **Rank Correlation Methods**. 2nd edition. By M. G. KENDALL. London: Charles Griffin and Co. 1955. Pp. 196. 36s.

2. **Manpower Shortage and the Fall of the Roman Empire**. By A. E. R. BOAK. U.S.A.: University of Michigan Press. 1955. Pp. 169. $4.50.

3. **The Foundations of Statistics**. By L. J. SAVAGE. U.S.A.: John Wiley and Sons; London: Chapman and Hall. 1954. Pp. xv + 294. 48s.

4. **Statistics in Research.** By B. OSTLE. U.S.A.: Iowa State College. 1954. Pp. xiv + 487. $6.95.

5. **Annual Epidemiological and Vital Statistics for 1952.** Switzerland: World Health Organization, Geneva. 1955. Pp. x + 533. 50s.

Publications of the U.S. Department of Commerce, National Bureau of Standards

6. **Tables of Sine and Cosine Integrals for arguments from 10 to 100.** Applied Mathematics Series, 32. 1954. Pp. xv + 187. $2.25.

7. **Tables of the Gamma Function for Complex Argument.** Applied Mathematics Series, 34, 1954. Pp. xvi + 105. $2.00.

8. **Tables of Functions and of Zeros of Functions.** Applied Mathematics Series, 37. 1954. Pp. ix + 211. $2.25.

9. **Contributions to the Solution of Systems of Linear Equations and the Determination of Eigen-values.** Edited by OLGA TAUSSKEY. Applied Mathematics Series, 39. 1954. Pp. 139. $2.00.

10. **Tables of the Error Function and its derivatives.** Applied Mathematics Series, 41. 1954. Pp. xi + 302. $3.25.